1999

Enhancing chromatographic separations of recombinant proteins from canola extracts by genetic design and characterization of protein binding regions

Chenming Zhang
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

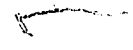
In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600



Enhancing chromatographic separations of recombinant proteins from canola extracts by genetic design and characterization of protein binding regions

by

Chenming Zhang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemical Engineering

Major Professor: Charles E. Glatz

Iowa State University

Ames, Iowa

1999

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Chenming Zhang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

To my wife Yuenan Chen and my parents Qiang Zhang and Chaolian Wang
for their encouragement and support

TABLE OF CONTENTS

ABSTRACT

Canola was studied as an alternative source to traditional microbial systems as a recombinant protein production host. Various chromatographic methods were used in this work to systematically characterize the native canola protein elution profiles, and different genetically engineered proteins were selected to explore the opportunities presented for effective protein recovery.

Native canola protein was eluted into two major peaks in linear salt gradient elution on cation-exchange chromatography. T4 lysozyme and its mutants, by both point mutation and fusion, were used as model proteins to investigate the charge effect on protein elution. It was found that the single mutant T4 lysozyme was eluted in the low canola background valley between the two major canola protein peaks under the experiment conditions, and the purity of collected lysozyme was more than 90%.

Canola protein elution profile presented three possible target sites for selective recombinant protein purification on a strong anion exchanger. β-Glucuronidase (GUS) and fusions were used as the model proteins. Wild-type GUS was found to be eluted roughly at the first target site with the least eluent salt concentration, and GUS fusions with more than 10 aspartates in the tail were moved to the second site (higher salt concentration) with lower canola protein background. The third site with the highest salt elution was out of reach. The enrichment factor of fusion GUS at the second site was three to four times higher than that of wild-type GUS.

Stoichiometric displacement model was used to characterize the series of GUS proteins (wild-type GUS and its fusions) under isocratic elution. It was found that the fusion

with 15 aspartates did not follow the trend of change in protein specific parameters, Z and I, as the number of aspartates was less than 10. The Z and I values were used in an equation and lumped dispersion model to predict the protein elution under various gradient slopes. Both methods could reasonably well predict the protein retention and the influence of gradient change on the protein elution, and the simulation could also pick up the shape of the protein peaks.

Immobilized metal affinity chromatography (IMAC) was also used to explore the possibility of purifying proteins with poly-histidine fusions. GUS-$(his)_6$, as our model protein, could be purified to almost homogeneous purity on Co^{2+} columns with iminodiacetate (IDA) and nitrilotriacetate (NTA) as the immobilization ligand. The recognition of metal ions on protein surface histidine distribution was found following the order $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$. The binding mechanism was proposed to describe the interaction between poly-his tagged protein with immobilized metal ions when IDA and NTA were used as the chelating ligands.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

Transgenic plants are being studied as potential recombinant protein production hosts. This work was undertaken to investigate the possibility of using canola for recombinant protein production. Native canola proteins were characterized by various chromatographic methods, including cation-exchange, anion-exchange, and immobilized metal affinity chromatography (IMAC) to assess the applicability of these methods for purification proteins expressed in canola. Additionally, genetically engineered proteins may enhance protein recovery, and appropriate model proteins were chosen to investigate this possibility. In this regard, the general ability to predict the elution of fusion proteins from anion-exchange chromatography was also studied. The main objectives of this work are outlined in the following paragraphs along with the specific questions addressed.

The first objective was to characterize the native canola protein elution on cation-exchange, anion-exchange, and immobilized metal affinity chromatography, and to seek means of genetically modifying recombinant proteins to be most easily separated from these canola proteins. The model proteins chosen were T4 lysozyme and β-Glucuronidase (GUS). T4 lysozyme was genetically modified by both point mutation and fusion. It carries positive charge at neutral pH and was studied by cation-exchange chromatography. GUS was modified only by fusions. For anion-exchange chromatography, in addition to the wild-type GUS, three fusions with 5, 10, or 15 aspartates in the fusion tail were studied. For IMAC, GUS with a $(his)_6$ tail (GUSH6) was studied on columns with four different metal ions (Cu^{2+},

Ni^{2+}, Zn^{2+}, Co^{2+}) immobilized by two different chelating ligands (iminodiacetate, IDA, and nitrilotriacetate, NTA). The specific questions to be answered are:

1. What does the native canola protein elution profile look like under linear gradient elution for each chromatographic method?
2. Does canola present any opportunity for enhanced protein purification?
3. Are GUS and its fusions stable in canola extracts?
4. For ion-exchange chromatography, what is the effect of protein charge on elution when loaded with canola extract? For IMAC, what is the effect of different metal ions and chelating ligands on GUSH6 elution?
5. Protein of which charge is most effectively purified from canola extract by a single step ion-exchange chromatography, and which metal ion and chelating ligand provide the best opportunity for GUSH6 recovery?
6. For T4 lysozyme, is there any difference between altering charge by point mutation or by fusion for purification from canola extract?
7. For IMAC, is there a way to interpret the affinity behavior of GUSH6 on different metal ions immobilized by different chelating ligands?

The second objective of this work was to mathematically simulate the retention of the series of genetically engineered fusion proteins in linear gradient elution. Two methods were used: an analytical equation and a thermodynamic model considering the axial dispersion effect during chromatographic process (lumped dispersion model). GUS and its aspartate fusions were first characterized by the parameters of the stoichiometric displacement model (SDM), and these parameters were then used to predict elution. The questions to be answered are:

1. How do the fusion tails influence the protein-binding parameters in SDM?
2. How well can the lumped dispersion model simulate the protein elution in linear gradient elution chromatography?
3. How well can the simpler analytical equation predict the protein elution for different gradient elution profiles?

Dissertation Organization

The dissertation contains five main chapters. The first chapter provides basic information on transgenic plants and the fundamental theories of chromatography used in this work.

The other four chapters present the experimental work performed to fulfill the objectives raised above, and they are presented in a journal paper format. The first paper (chapter 2) focuses on T4 lysozyme purification from canola extract by cation-exchange chromatography. T4 lysozyme mutants with different charges modified by either point mutation or fusion were spiked into canola extract to investigate the possibility of moving the protein to a favorable site for enhanced recovery. The second paper (chapter 3) follows the thoughts of the first one but uses anion-exchange chromatography with a negatively charged protein. The third paper (chapter 4) studies the ability to predict and simulate fusion protein elution from anion-exchange chromatography using two equilibrium models. A stoichiometric displacement model was used to characterize the fusion protein retention on a selected anion-exchanger to provide more understanding on the binding capability of proteins with different length of fusion tails. The SDM parameters were used in subsequent protein elution prediction and simulation. The fourth paper (chapter 5) follows the approach of the

first two papers in exploring the use of histidine-tagged recombinant proteins for enhanced recovery from canola extracts. Different metal ions on two different chelating agents were studied to determine the best purification method. The binding mechanisms of proteins on different chelating resins were proposed.

Two appendices are included. The first gives the amino acid sequences of GUS and its fusions. The second appendix includes the estimated net charge of GUS and its fusions at different pH.

Literature Review

Plants as recombinant protein production hosts

The emergence of recombinant DNA technology since the 1970's has allowed the production of large quantities of scarce and novel proteins having enormous medical, agriculture, and industrial impact [1]. Sources for foreign protein (recombinant protein) production have been mainly microorganisms such as *E. coli* (prokaryote) and yeast (eukaryote) but also animal cells and even transgenic animals. However, all these systems have shortcomings. Microbial systems do not provide accurate post-translational modification of eukaryotic proteins, and bacterial fermentation, which is still expensive and difficult to scale-up [2,3], often produces proteins as insoluble aggregates that are difficult to refold to their native, biologically active form. Using transgenic animals to produce heterologous proteins has raised social, ethical, and legal concerns [1, 4]. Transgenic plants, on the other hand, provide attractive and cost-effective alternatives to the above-mentioned systems for recombinant protein production. As eukaryotes, transgenic plants have the

capability for post-translational processing of proteins of eukaryotic origin (plant or animal) [5, 6].

The other advantages of using plants as recombinant protein production hosts include the ease of scale-up (by just increasing the planted acreage), and established methods of harvesting, processing, transporting, and storing for extended long period of time [5]. In addition, for vaccine production from transgenic plants, the fear of animal virus contamination could be eliminated, because plant viruses do not infect humans [7].

Plants have been used mainly as food protein and oil sources, but recent developments in plant genetic engineering have made plants commercially attractive for recombinant protein production for pharmaceuticals [4, 7, 8], agriculture [6, 9, 10], food [4], and commodities [11]. Since the late 1980's, many pharmaceutical proteins and peptides and industrial proteins have been successfully expressed in various plants (Table 1.1). The expressed proteins or peptides range from 0.6 kDa to as large as 80 kDa per subunit, and the original sources of expressed proteins cover a wide range of species, including human beings. The highest accumulation of a protein, phytase expressed in tobacco, reached 14% of the total soluble protein [36].

While the accumulation level of a recombinant protein is critical, the downstream processing is equally important. The downstream processing from plant biomass is generally considered to be difficult and expensive because of the large amount of total biomass to be handled [5, 6, 37]. However, concentrating and purifying the recombinant protein at an early stage would reduce costs [38]. Genetic engineering provides a possible way to accomplish this, and purification fusions could have particular impact.

Table 1.1. Examples of proteins and polypeptides produced in transgenic plants (modified from [5]).

Protein	Origin of gene(s)	Molecular weight (kDa/subunit)	Host	Expression level	Reference
α-amylase	*Bacillus licheniformis*	55.2	Tobacco,	0.3% of soluble leaf protein	12
			alfalfa	Not available	13
α-trichosantin	Chinese medicinal	27	N.	2% of total soluble protein	14
	plant		benthamiana[1]		
Avidin[2]	Chicken	16.8	Corn	6% of soluble seed protein	15
(1-3,1-4)-β-	*Trichoderma reesei*,	24	Barley cells	0.4% of extracellular protein	16, 17
glucanase	hybrid of two *Bacillus*				
	species				
β-glucuronidase	*E. coli*	68	Alfalfa,	Not available	13
			Corn	0.7% of soluble protein	18
Chymosin	Calf	30	Tobacco,	0.1-0.5% of soluble protein	19
			potato		
Enkephalin	Human	~1[3]	Canola	~0.02% of soluble seed protein	20, 21, 22

Table 1.1. (Continued)

Protein	Origin of gene(s)	Molecular weight (kDa/subunit)	Host	Expression level	Reference
Epidermal growth factor	Human	6	Tobacco	0.001% of soluble protein	23
Erythropoietin[4]	Human	~37	Tobacco	0.0026% of soluble protein	24
Growth hormone	Trout	21	Tobacco	0.1% of soluble leaf protein	25
Hirudin	Hirudo medicinalis	11	Canola	1% of seed weight	26
Lysozyme	Chicken	14.4	Tobacco	0.003% of fresh leaf tissue	27
Malarial epitopes	*Plasmodium*	0.6	Tobacco	0.4-0.8% of virion weight	28
Manganese-dependent lignin peroxidase	*Phanerochaete chrysosporium*	42	Alfalfa	Not available	13, 29
Phytase	*Aspergillus niger*	80	Tobacco	14.4% of soluble leaf protein	30
Ricin	Castor bean	34	Tobacco	0.25% of soluble leaf protein	31
Serum albumin	Human	66.5	Potato	0.02% of soluble leaf protein	32

Table 1.1. (Continued)

Protein	Origin of gene(s)	Molecular weight (kDa/subunit)	Host	Expression level	Reference
Xylanase	*Clostridium thermocellum*, *Cryptococcus albidus*	48[5]	Tobacco	4.1% of soluble leaf proteins	33, 34

1. Nicotiana benthamiana.

2. First commercial product produced by a plant system.

3. The molecular weight is determined from pig brain extract [22].

4. 37 (35-39) kDa is the glycosylated Erythropoietin. The deglycosylated protein has a molecular weight of 18-20 kDa. This is also the protein produced by Amgen from microbial fermentation (~250 gram/year) and sold at $1 million/gram [35].

5. A deleted, but active xylanase from *Clostridium thermocellum* has a molecular weight of 37 kDa [33].

Canola as a recombinant protein production host and its native protein properties

Canola as a recombinant protein production host

Tobacco has been routinely used as one of the model plants to test recombinant gene expression [39]. However, the high level of toxic alkaloids in its leaves demands careful purification of recombinant proteins in downstream processing, especially for immunotherapeutic and pharmaceutical proteins [7, 38]. In addition, seeds or tubers are preferred to green tissues as protein storage organs for recombinant protein production, because seeds and tubers are natural storage organs and allow long periods of storage without protein degradation. Hence, the necessity for green tissue plants of having a processing facility adjacent to the harvesting location could be eliminated [5, 13, 37]. Crops, such as potatoes and canola, have been the focus of attention for recombinant protein production because they are already used in producing the primary non-proteinaceous products - starch and oil [6]. Canola is one of the several crops for which reliable transformation systems have been developed and was predicted to be one of the first to be used as a protein factory [6], and it is chosen as our model system to explore its potential as a recombinant protein production host.

Canola, sometime also called rapeseed, is the third largest source of vegetable oil behind soybean and palm [40]. Its advantages have led to efforts to produce several recombinant proteins [20, 21, 26]. Even though the non-protein compounds, glucosinolates and their breakdown products, phenolics and phytic acids, have slowed the utilization of native canola proteins [41], the drawbacks can be minimized when the recombinant protein is to be purified.

Native canola protein properties

Though the protein content depends on variety and planting conditions, canola seed generally has 33-39% storage protein in deffated meals. The protein is mainly composed of a high-molecular weight (12S) and a low-molecular weight (1.7S or 2S) protein fractions [41]. The relative content of 12S and 2S protein varies with variety [42]. Table 1.2 shows the properties of these two protein groups.

The 12S protein has been reported to dissociate at low ionic strength ($\mu < 0.5$) into 7S components (molecular weight ~ 150 kDa) and at low or extremely high pH and in the

Table 1.2. Properties of 12S and 2S proteins in canola.

Protein group*	12S**	2S
Name	Cruciferin, Brassin	Napin, Conbrassin
Molecular weight (kDa)	300 ± 10	12 – 17
Isoelectric point	7.2	≥ 10
Molecular shape	Oblate ellipsoid of revolution	linear
Number of subunits	6	1
Secondary structure (%)	α-helix 10; β-sheet 50	α-helix 40-46; β-sheet 12
Reference	43 – 47	41, 48

* Each group of protein consists more than one protein. However, all the proteins in the same group have similar characteristics listed in this table.

** S is the sedimentation coefficient (Svedbergs), which is a measure of the protein molecular mass.

presence of urea further dissociate into smaller (2-3S) components [45, 49-51] (Fig. 1.1). Similar to other plant seed 11/12S storage proteins [43], extracted 12S protein can react with the phytic acid present in canola [52-54] and form insoluble and acidic complexes with pH in the range of 3-5 [43, 55]. Besides the above mentioned two groups of proteins, there are also some proteins with intermediate molecular weights of 50 kDa to 75 kDa with isoelectric points close to 8 [56].

Choices of purification methods

Choice of protein purification methods depends on a combination of many factors, which mainly include the characteristics of the target protein, the protein source, the purity requirement for the protein, the capability of a technique in handling high protein load, the economic feasibility of the technique, and ease of scale-up. Table 1.3 shows how method choice has changed since 1980.

Ion exchange chromatography is chosen as one of the means to carry out our research because of its popularity in plant protein separation and ability of handling large protein loads. A fast emerging adsorption technique, immobilized metal affinity chromatography (IMAC), will also be studied in this work. All the detailed separation mechanisms will be discussed below.

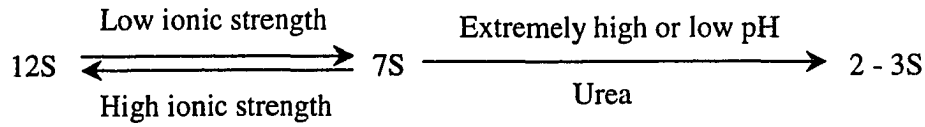


Figure 1.1. Schematic illustration of 12S protein dissociation under different conditions [51].

Table 1.3. The trend of using various methods in plant protein purification [57].

Method	Principle	Trend after 1980
Precipitation	salting out, isoelectric precipitation, electrostatic complex formation, etc	decrease
Ion-exchange chromatography	electrostatic interaction	
AEC		increase
CEC		decrease
Affinity chromatography		
Bio-ligand	biological	increase
Dye-ligand	various	increase
Gel filtration	molecular size	static
Hydrophobic interaction Chromatography	hydrophobic	increase
FPLC/HPLC*	various	increase

* FPLC: Fast performance liquid chromatography, or fast protein liquid chromatography.
HPLC: High performance liquid chromatography.

Ion exchange chromatography

Separation mechanism

Ion exchange chromatography is one of the most widely used techniques in protein purification [58]. The choice between cation and anion exchange chromatography depends on the charge that the target protein carries. The charge of a protein, with an amphoteric nature, depends on the protein's isoelectric point (pI, the pH at which the protein's net charge

is zero) (Fig. 1.2) and can be changed by adjusting the system pH. AEC is used when the protein carries net negative charges (pH > pI), and CEC is used when the protein carries net positive charges (pH < pI).

Ion exchange chromatography is a process of separating components by passing a sample solution through a column so that the components adsorb to the absorbent surface by electrostatic interaction to different extents based on their characteristics, and thus elute at different time from the column. Figure 1.3 is a conceptual model of the transport processes involved in binding to the adsorbent of a component added to the mobile phase. Different models have been proposed to describe the protein adsorption and predict the elution of proteins from an IEC column [58-67].

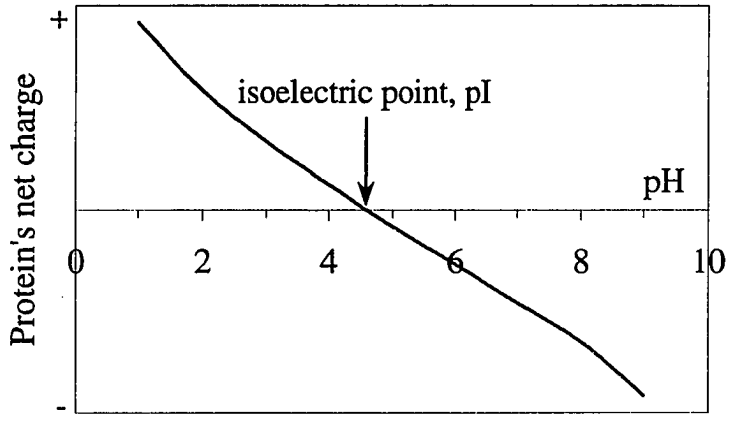


Figure 1.2. The relationship between the charge a protein carries and the system pH.

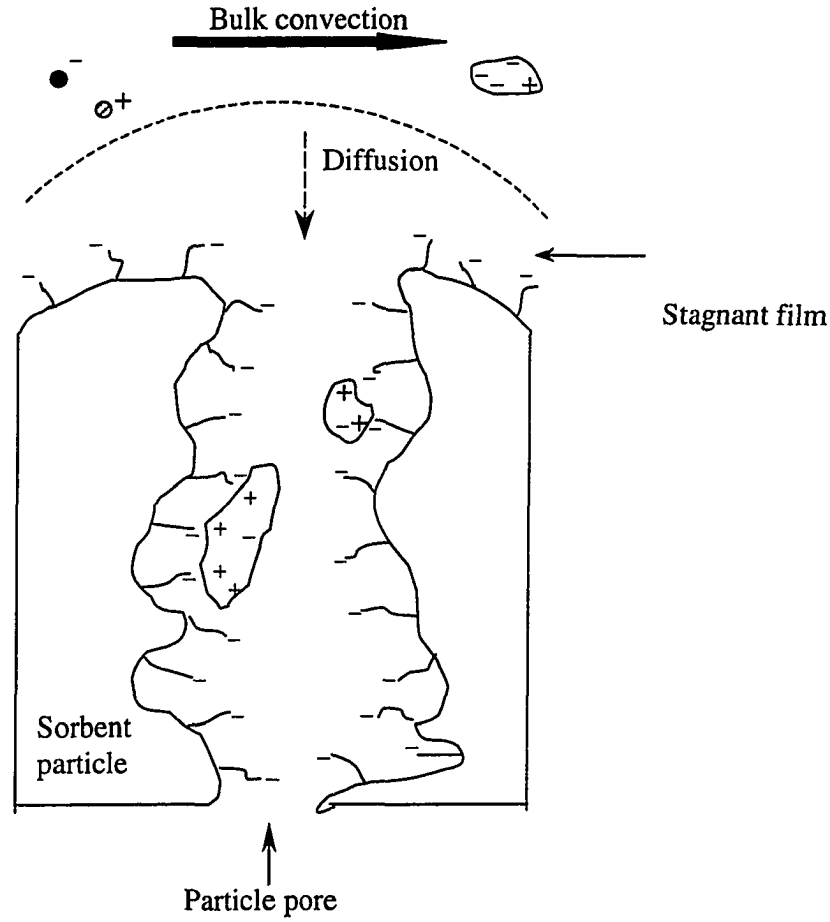


Figure 1.3. Schematic illustration of an ion-exchange process (cation exchange here).

One widely-used model, the stoichiometric displacement model (SDM), is based on the mass action law [58, 60, 65-68]. Protein adsorption is treated with the stoichiometry of the reaction:

$$P_0 + Z \cdot D_b \Leftrightarrow P_b + Z \cdot D \tag{1}$$

where a free protein (P_0) in solution will displace multiple (Z) bound ions (D_b) when it binds (P_b) to the ion-exchange sorbent surface. The stoichiometry of the displacement process is determined by the number (Z) of ions that are required to displace the adsorbed protein from the sorbent. The reaction equilibrium constant K_e can be expressed as

$$K_e = \frac{[P_b][D]^Z}{[P_0][D_b]^Z} \tag{2}$$

For low protein coverage, the protein retention (as the capacity factor k') can be related to displacing ion concentration (D) by the following equation [67, 68]

$$k' = \frac{I}{[D]^Z} \tag{3}$$

where I is a protein specific constant. Its value describes the affinity of the protein for the stationary phase through available chromatographic contact regions along with any nonspecific interactions involved with protein retention [69], and it is given by the equation

$$I = K_e \varphi [\Lambda]^Z \tag{4}$$

where φ is the phase ratio (ratio of stationary and mobile phase volumes) and Λ is the ionic capacity of the resin.

The capacity factor, k', is also related to the retention time of a protein at retained and non-retained conditions, t_R, and t_0, respectively, by

$$k' = \frac{(t_R - t_0)}{t_0} \tag{5}$$

The linear form of eq. (3),

$$\log k' = \log I + Z\log[D]^{-1} \tag{6}$$

can be used to obtain the two protein specific parameters, Z and I, with the k' obtained from eq. (5) by measuring t_R under isocratic elution at series of D.

Protein retention during gradient elution

A protein's retention in a gradient elution can be predicted by the equation derived by Jandera and Churacek [70, 71] using protein specific parameters, Z and I, described above. In isocratic elution, the protein capacity factor is defined as,

$$k' = \frac{V_R'}{V_m} = \frac{V_s}{V_m} \cdot \frac{[P_b]}{[P_0]} \tag{7}$$

where, V_R' is the reduced protein retention volume, and it equals the difference between protein retention volume (V_R) and the total volume of the mobile phase in the column (V_m). The ratio of total volume of the stationary phase in column (V_s) to V_m is the phase ratio, φ, in Eq. (4). By combining eq. (7), (2), and (4) and assuming dilute condition, eq. (3) can be obtained.

During gradient elution, the k' value in the course of a differential volume change is

$$k' = \frac{dV}{dV_m} \tag{8}$$

where dV is a differential increase in the volume of the mobile phase delivered from the gradient-generating device on to the column. This elution increment causes a migration of the band of a sample compound in the column through a distance corresponding to a

differential part of the column void volume, dV_m. The total volume of the mobile phase delivered to the column from the beginning of the elution (V, the volume of eluate) is

$$V = V'_{R(g)} - V_d = V_{R(g)} - V_m - V_d \tag{9}$$

Where, $V'_{R(g)}$ ($V_{R(g)}$) is the reduced (non-reduced) retention volume of a sample component in gradient elution chromatography. V_d is the dead volume of the connecting tubing between the outlet of the gradient generating device and the top of the column.

The protein retention volume can be obtained by the integration of eq. (8),

$$\int_0^{V_m} dV_m = \int_0^{V'_{R(g)}} \frac{1}{k'} \cdot dV \tag{10}$$

The displacing ion concentration (D) as a function of elution time (t) can be described by the following equation,

$$D = (D_e^{\frac{1}{\chi}} + G' \cdot t)^{\chi} \tag{11}$$

Where D_e is the displacing ion concentration in the equilibrating buffer at the beginning of the gradient elution, and G' is the slope of the gradient described in time scale. χ is the shape of the concentration gradient. When $\chi = 1$, it is a linear gradient elution, and the concentration gradient is concave for $\chi > 1$ and convex for $\chi < 1$.

Furthermore, since $t = V/u$ during elution, where u is the flow rate of the mobile phase in ml/min, the displacing ion concentration can be described by total elution volume b

$$D = (D_e^{\frac{1}{\chi}} + G \cdot V)^{\chi} \tag{12}$$

where $G = G'/u$, is a measure of the slope of the displacing ion concentration change during gradient elution in volume scale.

Replace D in eq. (3) by eq. (12) and substitute the k' into eq. (10),

$$V_m = \frac{1}{I}\int_0^{V'_{R(g)}} (D_e^{\frac{1}{\chi}} + G \cdot V)^{\chi \cdot Z} dV \tag{13}$$

and the solution of this equation gives the retention volume of a protein in gradient elution chromatography.

$$V'_{R(g)} = \frac{1}{G}\left[G \cdot (\chi \cdot Z + 1) \cdot I \cdot V_m + D_e^{\frac{\chi \cdot Z+1}{\chi}} \right]^{\frac{1}{\chi \cdot Z+1}} - \frac{D_e^{\frac{1}{\chi}}}{G} \tag{14}$$

From eq. (9) and (12), the concentration of the displacing ion in the mobile phase which elutes the protein peak maximum in gradient elution, $D_{(g)}$, can be calculated by

$$D_{(g)} = \left[D_e^{\frac{1}{\chi}} + G \cdot (V'_{R(g)} - V_d) \right]^{\chi} \tag{15}$$

The eluting salt concentration as a function of protein parameters, Z and I, can be obtained by substituting eq. (14) into (15). This gives

$$D_{(g)} = \left\{ \left[G \cdot (\chi \cdot Z + 1) \cdot I \cdot V_m + D_e^{\frac{\chi \cdot Z+1}{\chi}} \right]^{\frac{1}{\chi \cdot Z+1}} - G \cdot V_d \right\}^{\chi} \tag{16}$$

Under linear gradient elution, when $\chi = 1$,

$$D_{(g)} = \left[G \cdot (Z + 1) \cdot I \cdot V_m + D_e^{Z+1} \right]^{\frac{1}{Z+1}} - G \cdot V_d \tag{17}$$

Eq. (17) can be further simplified when the concentration of displacing ion (e.g. Cl^- in AEC) in the mobile phase equals zero at the beginning of the gradient elution, i.e. D_e=0. It then gives,

$$D_{(g)} = \left[G \cdot (Z + 1) \cdot I \cdot V_m \right]^{\frac{1}{Z+1}} - G \cdot V_d \tag{18}$$

In eq. (18), except Z and I, other parameters, G, V_m, and V_d, are all directly measurable. Given a specific system (pH and column) under pre-chosen gradient slope, the above equation could be used to predict the salt concentration eluting a protein based on its Z and I values, which are obtained from isocratic elution chromatography experiments.

For a multi-component protein system (such as canola protein extract), if the loading is much smaller than the column capacity, eq. (17) or (18) is expected to be able to be applied to evaluate the protein retention in gradient elution.

There are also other ways to predict a protein's retention in gradient elution by using models which consider the mass transport processes described in Fig. 1.3. However, they often need more than just the protein specific parameters. Even though they are more likely to be accurate, the simplicity has to be sacrificed. For example, the lumped dispersion model is an equilibrium model that considers the axial dispersion of the solute in the mobile phase during a chromatographic process [72]. To use this model to simulate a protein's elution, it not only needs the two parameters discussed in SDM but also the protein's dispersion coefficient and some column specific parameters, such as column capacity and porosity [73]. In addition, the solution of the mass balance equation (partial differential equation) must be done numerically.

Immobilized metal affinity chromatography (IMAC)

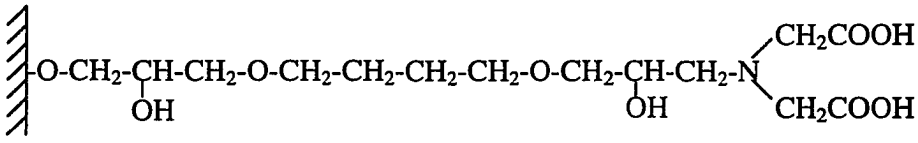
IMAC has become a popular technique in protein purification since introduced in 1975 by Porath [74]. Traditional affinity chromatography (using biospecific binding sites such as inhibitors or monoclonal antibodies) is a compromise among high selectivity, simplicity, and high cost. It is very protein specific and only used for meeting high purity requirements. Use of multiple steps of lower selectivity may be cost effective, but typically,

simplicity and yield are sacrificed. However, IMAC has shown ease of application to a whole class of proteins, and it is becoming one of the techniques of choice in protein separation, especially in recombinant protein purification [75]. Table 1.4 compares IMAC to bio-specific affinity chromatography .

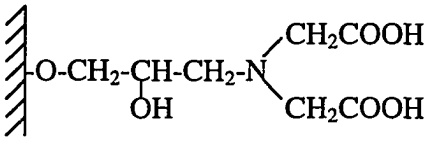
In IMAC, first-row transition metal ions (Ni^{2+}, Zn^{2+}, Cu^{2+}, Co^{2+}, and Fe^{3+}) are usually immobilized on a support (such as sepharose or agrose gel) by chelating through either iminodiacetic acid (IDA) or tris (carboxymethyl) ethylene-diamine (TED) (Fig. 1.4 and 1.5). However, IDA is more often used than TED, because the coupling of TED results in a stronger retention of metal (Me^{2+}) ions but weaker retention of a protein [77]. Several other chelating agents such as carboxymethylated aspartic acid (CM-ASP), tetraethylene pentamine (TEPA), carboxymethylated diamino succinic α,β acid (CM-DASA) [78], nitrilotriacetic acid (NTA) [79] and tris(2-aminoethyl)amine (Tren) [80], tetradentate

Table 1.4. Comparison of IMAC with biospecific affinity chromatography [76].

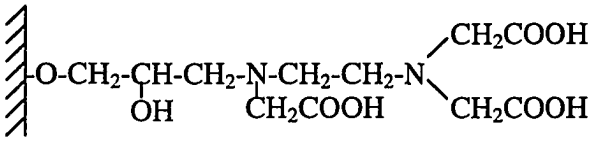
Property	Biospecific affinity	Metal-affinity
Ligand stability	Low	High
Protein loading	Low	High
Elution conditions	Often extreme	Mild
Ligand recovery after regeneration	Generally incomplete	Complete
Selectivity	High	Low-medium
Cost	High	Low



Chelating Sepharose 6B

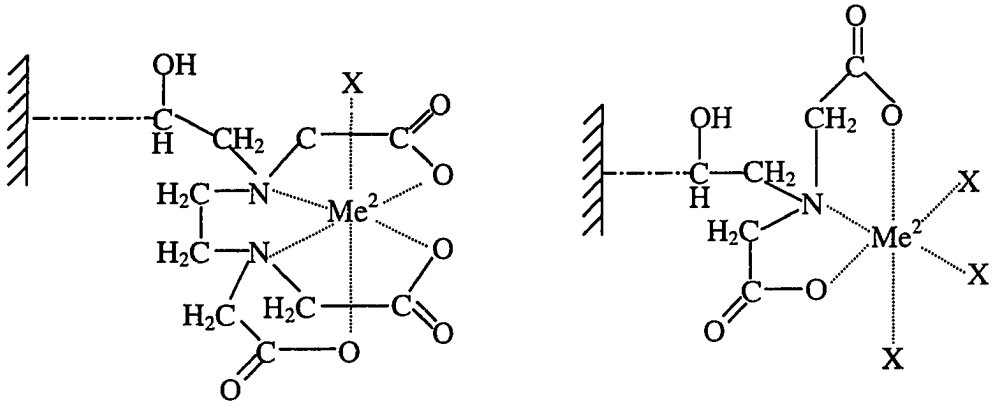


IDA-Sepharose 6B



TED-Sepharose 4B

Figure 1.4. Partial gel structures utilized in metal ion immobilization in IMAC [77].



(a) TED-Me^{2+} chelate (b) IDA-Me^{2+} chelate

Figure 1.5. Metal ion (Me^{2+}) immobilization on gel surface. TED-Me^{2+} chelate only leaves one available orbital for electron donors (a); meanwhile IDA-Me^{2+} chelate provides three available orbitals (b) [77].

chelators, have also been used, but IDA-gel and NTA-gel are the only commercially available gels for IMAC.

The interaction between a protein and immobilized metal ions is achieved by the electron donating side chains on the protein surface which bind to the metal ions by replacing water molecules [77]. Many functional group including amino acids such as Glu, Tyr, Cys, His, Arg, Lys, Asp, and Met, and the amino-terminal and backbone amide nitrogens can be involved in metal recognition, but fortunately, only histidines and cysteines on protein surface contribute significantly to protein retention [76]. Since reduced cysteines are rarely available on protein surface, histidine-metal interaction is the dominant force in protein retention. Furthermore, histidyl residues in different microenvironments on the protein surface have different affinity for metal ions because of differences in steric accessibility and the pK_a of the histidines. Other amino acids contribute to metal affinity by mainly indirectly affecting the pK_a and accessibility of histidine and solvation of the metal ion and histidine (Table 1.5) [76, 81]. At high enough pH, where the side chains of Lys, Arg, and Tyr are deprotonated, their contribution to protein-metal affinity becomes direct. The contribution of N-terminus to protein retention is small because the separation is often operated at pH ~ 7, where the N-terminus is protonated.

The distribution of histidine residues is also very important [76-78, 82] (Table 1.6). Protein-metal affinity has been shown to be directly proportional to the number of accessible histidines. The formation of multiply coordinated bonds can greatly increase the protein-metal binding strength and, therefore, enhance the selectivity of protein-metal interaction [82]. Furthermore, different histidine distributions on the protein surface show different affinities for immobilized metal ions (Table 1.7) [83]. In some cases (e.g. His, and HisHis),

Table 1.5. Relative contribution of different amino acid side chains and N-terminus of to protein-metal affinity in Cu^{2+}-, Ni^{2+}- and Zn^{2+}-IDA columns [1, 76, 81].

Functional group	pK_a	Contribution	Mechanism
His	6.04	++++	direct
Cys	8.37	++++	direct
Asp	3.90	-	indirect
Glu	4.07	-	indirect
Lys	10.54	+	indirect
Arg	12.48	+	indirect
Tyr	10.46	+	indirect
Trp	(Aromatic side chain)	+	indirect
Phe	(Aromatic side chain)	+	indirect
N-terminus	8	++	direct

even though the distribution of histidine residues will not facilitate the formation of multi-point chelating with one metal ion, their presence on a protein surface will still enhance its retention.

The typical IMAC metals are the first row transition metals Cu^{2+}, Zn^{2+}, Ni^{2+}, and Co^{2+}. The protein retention and stability constants for imidazole-metal ion complex on different metals follow the order $Cu^{2+} > Ni^{2+} > Zn^{2+} \sim Co^{2+}$ [76]. However, their presence at low concentration in protein solution can be potentially toxic when metal ion leakage happens during protein purification, whereas environmentally more tolerable metals, mainly alkaline earth and alkali metals, such as Ca^{2+} and K^{+}, have also been studied in order to

Table 1.6. Possible metal-chelation sites in α helix, β strand, and reverse β turn: (+) means chelation is possible; and (-) means chelation cannot occur. X = amino acid. (Histidine residues form a metal chelate with one single metal ion) [82].

Amino acid sequence	α helix	β strand	reverse β turn
His-His	-	-	-
His-X-His	-	+	-
His-X_2-His	-	-	+
His-X_3-His	+	-	-
His-X_4-His	-	-	-

Table 1.7. Distribution of surface histidine and immobilized metal ion recognition (IDA-Me^{2+}). Where (+) represents retention observed, and (-) represents no retention observed [83].

Sequence/ Me^{2+} *	Cu^{2+}	Ni^{2+}	Zn^{2+}	Co^{2+}
-His-	+	-	-	-
-His$(X)_n$His-	+	+	-	-
-His $(X)_{2,3}$His- (α-helix)	+	+	+	+
-HisHis-	+	+	+	+

* 20 mM Phosphate buffer (1M NaCl), pH 7.0.

develop less toxic IMAC [84, 85]. Other metals, such as Al^{3+}, Ga^{3+}, In^{3+}, and Tl^{3+}, have also been documented for potential applications in protein purification with IMAC [86, 87]. On the other hand, should undesired metal leaching occur, metal ions in protein solution can be removed by a metal free chelating column following the immobilized metal separation column [77].

During protein adsorption, in order to take full advantage of the specific metal-histidine interaction, high salt concentration adsorption buffer (1M NaCl) (pH 6-8) is usually used to quench non-specific electrostatic interactions [77, 80]. To recover the adsorbed protein, one of three methods can be used: 1) protonation, 2) ligand exchange, and 3) chelate annihilation (Fig. 1. 6).

When electron donor groups on protein surface are protonated by applying lower buffer pH during elution, they no longer bind to the metal and the protein is displaced. However, the decreased pH can also cause possible protein activity losses, isoelectric precipitation of proteins, and even bleaching of a metal column (loss of metal ions).

Proteins are eluted at near neutral pH by competing electron donor solutes (for

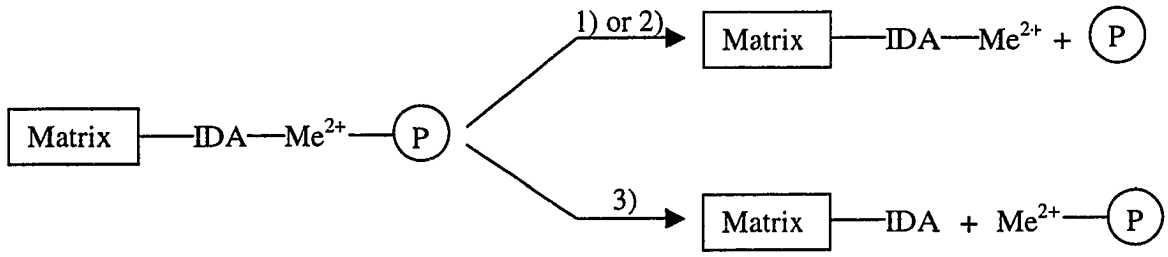


Figure 1.6. Schematic illustration of adsorbed protein (P) elution in IMAC by different desorption methods.

example, imidazole) for the ligand exchange mechanism. In this case, column pre-saturation and equilibrating is necessary.

Chelate annihilation occurs when the chelator-Me^{2+} complex itself is destroyed at low pH by a mild chelating agent (histidine) or a strong chelating agent (EDTA), and it results in the elution of protein-Me^{2+} complex [77].

Several models have been applied to describe protein retention in affinity chromatography. These include the Langmuir [88-90], Langmuir-Freundlich [91], Temkin [92, 93], and metal affinity interaction chromatography (MAIC) models [94, 95].

Both the Temkin and MAIC models take into consideration the multipoint interaction between protein and stationary phase (Fig. 1.7), which can well describe the increase in limiting capacity with increasing surface histidine content for what are otherwise identical proteins. However, none of the models describe how a protein with a poly-histidine tail would bind to the immobilized metal ions. Nonetheless, the MAIC model is useful in extracting some protein specific parameters such as the protein binding constant and the number of interaction sites.

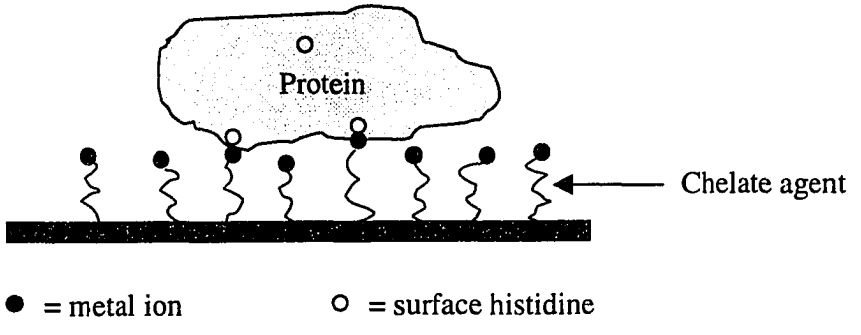


Figure 1.7. Schematic illustration of multiple site interaction during protein adsorption in IMAC.

Protein genetic engineering

Genetic engineering (recombinant DNA) techniques have long been used to facilitate purification of recombinant proteins with the modifications of the protein matched to the intended purification method [96]. Examples are summarized in Table 1.8. Both site-directed mutagenesis (point mutation) and introduction of an affinity tag by means of a fusion complement (fusion) have been used to introduce the modifications. Using point mutation, a specific amino acid residue on the protein sequence can be replaced by another amino acid. Using genetic fusions, a sequence of amino acids or a complete protein can be attached to either the C or N terminal of a target protein.

In order to perform directed point mutations, not only the primary sequence of a protein has to be known, but also its tertiary structure and functional domains in order to avoid inactivation. On the other hand, adding a fusion tail to a target protein is easier and has succeeded even in the absence of structural information [96, 112, 113]. Moreover, if necessary, the fusion can be removed by specific proteolytic cleavage to maintain the integrity of a target protein after purification. Various fusion tails, from an entire biologically functional protein to a few amino acids (polypeptide), have been used (Table 1.8). Small polypeptide tails, polyarginine and polyglutamic acid tails, have been used to facilitate the purification of target proteins with ion exchange chromatography [102, 105, 109, 110], and polyhistidine tails have been widely used to enhance the interaction of a target protein with immobilized metal ions in IMAC [97-99, 103, 107]. The advantage of using small fusion tails is that the tails usually do not affect the biological functioning of target proteins; hence, it is often not necessary to remove them.

Table 1.8. Examples of proteins genetically engineered for ease of purification.

Protein	Genetic engineering method	Specific modification	Purification method	Ref.
Citrate synthase	C/N-terminal fusion	His(5)*	IMAC	97
Dihydrofolate reductase	C/N-terminal fusion	His(2,3,4,5,6)	IMAC	98
Disulfide isomerase	C-terminal fusion	His(6)	IMAC	97
F_v fragment	C-terminal fusion	His(5)	IMAC	99
β-galactosidase	C-terminal fusion	Poly-aspartic acid	Extraction/Pcpt.	100, 101
			AEC	102
β-galactosidase	N-terminal fusion	His(6)	IMAC	103
Glucoamylase	C/N-terminal fusion	Poly-aspartic acid	Precipitation	104
hGH	N-terminal fusion	Poly-Glutamic acid	AEC	105
hGRF	Fusion	PhoS	Hydroxylapatite-AC	106
HIV-1 RT	C/N-terminal fusion	His(6)	IMAC	107
IGF-1	Fusion	Protein A	IgG-AC	108
T4 lysozyme	Point mutation	Lys to Glu	Extraction	100, 101
Urogastrone	C-terminal fusion	Polyarginine(5)	CEC	109, 110
Somatotropin	Point mutation	Amino acids to His	IMAC	111

* The number in the parentheses is the number of the specific amino acid in the fusion tail. HIV-1 RT, human immunodeficiency virus (HIV-1) reverse transcriptase. IGF-1, human insulin-like growth factor I. AC, affinity chromatography. hGRF, human growth-hormone-releasing factor. PhoS, phosphate-binding protein. HGH, human growth hormone. Pcpt., precipitation.

Like other protein production systems, the development of fusion proteins will also be of importance in improving the downstream processing of the target proteins in recombinant protein production from plant systems [6]. It is one of the goals of this work to provide the information on how fusions can enhance the recovery of a target protein from canola.

Acronyms

AEC: anion exchange chromatography

CEC: cation exchange chromatography

CM-DASA: carboxymethylated diamino succinic α,β acid (metal ion chelating agent in IMAC)

CM-ASP: carboxymethylated aspartic acid (metal ion chelating agent in IMAC)

HPLC: high performance liquid chromatography

IDA: iminodiacitate acid (metal ion chelating agent in IMAC)

IEC: ion exchange chromatography

IMAC: immobilized metal affinity chromatography

MAIC: metal affinity interaction chromatography model

NTA: nitrilotriacetic acid (metal ion chelating agent in IMAC)

SDM: stoichiometric displacement model for IEC

SEC: size exclusion chromatography

TED: tris (carboxymethyl) ethylene-diamine (metal ion chelating agent in IMAC)

TEPA: tetraethylene pentamine (metal ion chelating agent in IMAC)

Symbols

D: displacing ion concentration in mobile phase (M).

D_b: bound ion concentration in stationary phase.

D_e: the displacing ion concentration at the beginning of the gradient elution (M).

$D_{(g)}$: the concentration of the displacing ion in the mobile phase which elutes the protein peak in gradient elution.

$G = G'/u$: a measure of the slope of the displacing ion concentration change during gradient

G': the slope of the gradient described in time scale.

I: protein specific constant.

k': protein capacity factor.

K_e: protein adsorption equilibrium constant in IEC.

P_0: free protein concentration in solution (M).

P_b: bound protein concentration in stationary phase.

t: elution time (min).

t_0: retention time of a non-retained species in IEC under isocratic elution.

t_R: the retention time of a retained protein in IEC under isocratic elution.

u: the volumetric flow rate (ml/min).

V_m : the total volume of the mobile phase in a column.

V_R : non-reduced retention volume.

$V_R^{'} = V_R - V_m$: the reduced protein retention volume.

V_s : the total volume of the stationary phase in column.

V_d : the dead volume of the connecting tubing from the outlet of the gradient generating device to the top of the column.

Z: number of binding sites of a protein in SDM in IEC.

φ: the phase ration of a column (the ratio of stationary and mobile phase volumes).

χ : the shape of the concentration gradient.

Λ: the ionic capacity of the resin in a column.

References

[1] Voet, D. and Voet, J.G., Biochemistry (2nd ed.), John Wiley & Sons, Inc., 1995; pp906-909.

[2] Goosen, M.F.A., In *Insect Cell Culture Engineering*; Goosen, M.F.A., Baugulis, A., Faulkner, P., Eds.; Marcel Dekker, Inc.: New York, 1993; pp 1-16.

[3] Luckow, V.A. and Summers, M.D., Trends in the development of baculovirus expression vectors, *Bio/Technol.*, 6 (1988), 47-55.

[4] Krimsky, S. and Wrubel, R.P., In Agricultural biotechnology and the environment: Science, policy, and social issues, University of Illinois Press: Urbana, 1996; pp 99-113, 191-211.

[5] Kusnadi, A., Nikolov, Z.L. and Howard, J., Production of recombinant proteins in transgenic plants: Practical considerations, *Biotech. Bioeng.*, 56 (1997), 473-484.

[6] Whitelam, G.C., Cockburn, B., Gandecha, A.R. and Owen, M.R.L., Heterologous protein production in transgenic plants, *Biotechnol. Genet. Eng. Rev.*, 11 (1993), 1-29.

[7] Moffat, A.S., Exploring transgenic plants as a new vaccine source, *Science*, 268 (1995), 658-660.

[8] Mason, H.S. and Arntzen, C.J., Transgenic plants as vaccine production systems, *Trends Biotecnol.*, 13 (9) (1995), 388-392.

[9] Pen, J., Verwoerd, T.C., van Paridon, P.A., Beudeker, R.F., van den Elzen, P.J.M., Geerse, K., van der Klis, J. D., Versteegh, H.A.J., van Ooyen, A.J.J. and Hoekema, A., Phytase-containing transgenic seeds as a novel feed additive for improved phosphorus utilization, *Bio/Technol.*, 11 (1993), 811-814.

[10] Romano, C.P. and Klee, H., *Hormone manipulation in transgenic plants, In* Transgenic plants: Fundamentals and applications, Hiatt, A. Eds., Marcel Dekker, Inc.: New York, 1993; pp 23-36.

[11] Ponstein, A.S., Verwoerd, T.C. and Pen, J., Production of enzymes for industrial use, *Ann. NY Acad. Sci.*, 792 (1996), 91-98.

[12] Pen, J., Molendijk, L., Quax, W.J., Sijmons, P.C., van Ooyen, A.J.J., van den Elzen, P.J.M., Rietveld, K. and Hoekema, A., Production of active Bacillus licheniformis alpha-amylase in tobacco and its application in starch liquefaction, *Bio/Technol.*, 10 (1992), 292-296.

[13] Austin, S., Bingham, E.T., Koegel, R.G., Mathews, D.E., Shahan, M.N., Straub, R.J. and Burgess, R.R. An overview of a feasibility study for the production of industrial enzymes in transgenic alfalfa, *Ann. NY Acad. Sci.*, 721 (1994), 234-244.

[14] Kumagai, M.H., Turpen, T.H., Weinzettl, N., Della-Cioppa, G., Turpen, A.M., Donson, J., Hilf, M.E., Grantham, G.L., Dawson, W.O., Chow, T.P., Piatak Jr., M. and Grill,

L.K., Rapid, high-level expression of biologically active α-trichosanthin in transfected plants by an RNA viral vector, *Proc. Natl, Acad. Sci.* USA, 90 (1993), 427-430.

[15] Hood, E.E., Witcher, D.R., Maddock, S., Meyer, T., Baszcynski, C., Bailey, M., Flynn, P., Register, J., Marshall, L., Bond, D., Kullisek, E., Kusnadi, A., Evangelista, R., Nikolov, Z., Wooge, C., Mehigh, R.J., Hernan, R., Kappel, W.K., Ritland, D., Li, C.-P. and Howard, J.A., Commercial production of Avidin from transgenic maize: Characterization of transformant, production, processing, extraction and purification, *Mol. Breed.*, 3 (1997), 291-306.

[16] Aspegren, K. Mannonen, L., Ritala, A., Puupponen-Pimia, R., Kurten, U., Salmenkallio-Marttila, M., Kauppinen, V. and Teeri, T.H., Secretion of a heat-stable fungal β-glucanase from transgenic, suspension-cultures barley cells, *Mol. Breed.*, 1 (1995), 91-99.

[17] Phillipson, B.A., Expression of hybrid (1-3, 1-4) -β-glucanase in barley protoplasts, *Plant Sci.*, 91 (1993), 195-206.

[18] Kusnadi, A.R., Hood, E.E., Witcher, D.R., Howard, J.A. and Nikolov, Z.L., Production and purification of two recombinant proteins from transgenic corn, to be published.

[19] Wilmitzer, L., Sonnewald, U., Roeber, M. and Carlsen, S.K., Transgenic plants expressing genes for industrial enzymes, International Patent, 1992, WO 92/01042.

[20] Krebbers, E. and van de Kerckhove, J., Production of peptides in plant seeds, *Trends Biotechnol.*, 8 (1990), 1-3.

[21] Vandekerckhove, J., Damme, J.V., Lijsebettens, M.V., Botterman, J., Block, M.D., Vandewiele, M., Clercq, A.D., Leemans, J., Montagu, M.V. and Krebbers, E.,

Enkephalins produced in transgenic plants using modified 2S seed storage proteins, *Bio/Technol.*, 7 (1989), 929-932.

[22] Hughes, J., Smith, T., Morgan, Morgan, B. and Fothergill, L., Purification and properties of enkephalin – the possible endogenous ligand for the morphine receptor, *Life Sci.*, 16 (1975), 1753-1758.

[23] Higo, K.-I., Saito, Y. and Higo, H., Expression of a chemically synthesized gene for human epidermal growth factor under the control of cauliflower mosaic virus 35S promoter in transgenic tobacco, *Biosci. Biotechnol. Biochem.*, 57 (1993), 1477-1481.

[24] Matsumoto, S., Ishii, A., Ikura, K., Ueda, M. and Saski, R., Expression of human erythropoietin in cultured tobacco cells, *Biosci. Biotechnol. Biochem.*, 57 (1993), 1249-1252.

[25] Bosch, D., Smal, J. and Krebbers, E., A trout growth hormone is expressed, correctly folded and partially glycosylated in the leaves but not the seeds of transgenic plants, *Transgenic Res.*, 3 (1994), 304-310.

[26] Parmenter, D.L., Boothe, J.G., van Rooijen, G.J.H., Yeung, E.C. and Moloney, M.M., Reduction of biologically active hirudin in plant seeds using oleosin partitioning, *Plant Mol. Biol.*, 29 (1995), 1167-1180.

[27] Trudel, J., Potvin, C. and Asselin, A., Expression of active hen egg white lysozyme in transgenic tobacco, *Plant Sci.*, 87 (1992), 55-67.

[28] Turpen, T.H., Reinl, S.J., Charoenvit, Y., Hoffman, S.L., Fallarme, V. and Grill, L.K., Malarial epitopees expressed on the surface of recombinant tobacco mosaic virus, *Bio/Technol.*, 13 (1995), 53-57.

[29] Tien, M. and Kirk, T.K., Lignin-degrading enzyme from *Phanerocheate chrysosporium*: Purification, characterization, and catalytic properties of a unique H_2O_2-requiring oxygenase, *Proc. Natl. Acad. Sci.* USA, 81 (1984), 2280-2284.

[30] Verwoerd, T.C., van Paridon, P.A., van Ooyen, A.J.J., van Lent, J.W.M., Hoekema, A. and Pen, J., Stable accumulation of Aspergillus niger phytase in transgenic tobacco leaves, *Plant Physiol.*, 109 (1995), 1199-1205.

[31] Sehnke, P.C., Pedrosa, L., Paul, A.L., Frankel, A.E. and Ferl, R.J., Expression of active processed ricin in transgenic tobacco, *J. Biol. Chem.*, 269 (1994), 22473-22476.

[32] Sijmons, P.C., Dekker, B.M., Schrammeijier, B., Verwoerd, T.C., van den Elzen, P.J.M. and Hoekema, A., Production of correctly processed human serum albumin in transgenic plants, *Bio/Technol.*, 8 (1990), 217-221.

[33] Herbers, K., Wilke, I. and Sonnewald, U., A thermostable xylanase from *Clostridium thermocellum* expressed at high levels in the apoplast of transgenic tobacco has no detrimental effects and is easily purified, *Bio/Technol.*, 13 (1995), 63-66.

[34] Laliberte, J.-F., Nicolas, O., Durand, S. and Morsoli, R., The xylanase introns from *Cryptococcus albidus* are accurately spliced in transgenic tobacco plants, *Plant Mol. Biol.*, 18 (1992), 447-451.

[35] Humphrey, A., Engineering challenges in the application of rDNA technology: yesterday, today, and tomorrow, *Ann. NY Acad. Sci.*, 721 (1994), 1-11.

[36] Verwoerd, T.C., van Paridon, P.A., van Ooyen, A.J.J., van Lent, J.W.M., Hoekema, A. and Pen, J., Stable accumulation of *Aspergillus niger* phytase in transgenic tobacco leaves, *Plant Physiol.*, 109 (1995), 1199-1205.

[37] Whitelam, G.C., The production of recombinant proteins in plants, *J. Sci. Food Agr.*, 68 (1995), 1-9.

[38] Goddijn, O.J.M. and Pen, J., Plants as bioreactors, *Trends Biotechnol.*, 13 (1995), 379-387.

[39] Uchimiya, H., Handa, T. and Brar, D.S., Transgenic plants, *J. Biotechnol.*, 12 (1989), 1-20.

[40] Oil crops yearbook, July 1995, USDA.

[41] Schwenke, K.D., Rapeseed proteins, In *New and developing sources of food proteins*, Hudson, B.J.F., Eds, Chapman & Hall, 1994; pp 281-306.

[42] Raab, B., Leman, H., Schwenke, K.D. and Kozlowska, H., Comparative study of the protein patterns of some rapeseed (Brassica napus L.) varieties by means of polyacrylamide gel electrophoresis and high-performance liquid chromatography, *Nahrung*, 36 (1992), 239-247.

[43] Schwenke, K.D., Raab, B., Linow, K.-J., Pahtz, W. and Uhlig, J., Isolation of the 12 S globulin from rapeseed (Brassica napus L.) and characterization as a "neutral" protein, *Nahrung*, 25 (1981), 271-280.

[44] Schwenke, K.D., Schultz, M., Linow, K.-J., Gast, K. and Zirwer, D., Hydrodynamic and quasi-elastic light scattering studies on the 12S globulin from rapeseed, *Int. J. Peptide Protein Res.*, 16 (1980), 12-18.

[45] Plietz, P., Damaschun, G., Muller, J.J., and Schwenke, K.D., The structure of 11-S globulin from sunflower and rape seed: A small-angle X-ray scattering study, *Eur. J. Biochem.*, 130 (1983), 315-320.

[46] Schwenke, K.D., Linow, K.-J. and Zirwer, D., Modification of the oligomeric structure of 11 S globulin from sunflower (Helianthus annuus L.) and rape (Brassica napus L.) seeds by succinylation, *Nahrung*, 30 (1986), 263-270.

[47] Zirwer, D., Gast, K., Welfle, H., Schlesier, B., and Schwenke, K.D., Secondary structure of globulins from plant seeds: a re-evaluation from circular dichroism measurements, *Int. J. Biol. Macromol.*, 7 (1985), 105-108.

[48] Lonnerdal, B. and Janson, J.-C., Studies on Brassica seed proteins, *Biochim. Biophys. Acta*, 278 (1972), 175-183.

[49] Bhatty, R.S., McKenzie, S.L., and Finlayson, A.J., The proteins of rapeseed (Brassica napus L.) soluble in sale solution, *Can. J. Biochem.*, 46 (1968), 1191-1197.

[50] Schwenke, K.D. and Linow, K.-J., A reversible dissociation of the 12 S globulin from rapeseed (Brassica napus L.) depending on ionic strength, *Nahrung*, 26 (1982), K5-K6.

[51] Schwenke, K.D., Raab, B., Plietz, P and Damaschun, G., The structure of the 12S globulin from rapeseed (Brassica napus L.), *Nahrung*, 27 (1983), 165-175.

[52] Siy, R.D. and Talbot, F.D.F., Preparation of low-phytate rapeseed protein by ultrafiltration: I. The aqueous extraction of phytate from deoiled rapeseed meals, *J. Am. Oil. Chemists Soc.*, 59 (1982), 191-194.

[53] Gillberg, L. and Tornell, B., Preparation of rapeseed protein isolates: Dissolution and precipitation behavior of rapeseed proteins, *J. Food Sci.*, 41 (1976), 1063-1069.

[54] Schwenke, K.D., Mothes, R., Gast, K. and Zirwer, D., Rapeseed protein – polyanion interactions, Protein Interact., Symp. 201st Annu. Meet. Am. Chem. Soc., 1991 (Pub. 1992).

[55] Schwenke, K.D., Elkowicz, K. and Kozlowska, H., Solubility of the 12S globulin from rapeseed effected by phytic acid, *Nahrung*, 23 (1979), 967-970.

[56] Lonnerdal, B., Gillberg, L. and Tornell, B., Preparation of rapeseed protein isolates: A study of rapeseed protein isolates by molecular sieve chromatography, *J. Food Sci.*, 42 (1977), 75-78.

[57] Jervis, L. and Pierpoint, W.S., Purification technologies for plant proteins, *J. Biotechnol.*, 11 (1989), 161-198.

[58] Yamamoto, S., Nakanishi, K. and Matsuno, R., Eds., Ion-exchange chromatography of proteins, Chromatographic science series, 43, Marcel Dekker, Inc.: New York, 1988.

[59] Berninger, J.A., Whitley, R.D., Zhang, X. and Wang, N.-H. L., A versatile model for simulation of reaction and nonequilibrium dynamics in multicomponent fixed-bed adsorption process, *Comput. Chem. Eng.*, 15 (1991), 749-768.

[60] Boardman, N.K. and Partridge, S.M., Separation of neutral proteins on ion-exchange resins, *Biochem. J.*, 59 (1955), 543-552.

[61] Hasnat, A. and Juvekar, V.A., Ion-exchange kinetics: Heterogeneous resin-phase model, *AIChE J.*, 42 (1996), 161-175.

[62] Gallant, S.R., Vunnum, S. and Cramer, S.M., Modeling gradient elution of proteins in ion-exchange chromatography, *AIChE J.*, 42 (1996), 2511-2520.

[63] Velayudhan, A. and Horvath, C., Adsorption and ion-exchange isotherms in preparative chromatography, *J. Chromatogr. A*, 663 (1994), 1-10.

[64] Brooks, C.A. and Cramer, S.M., Steric mass-action ion-exchange: displacement profiles and induced salt gradients, *AIChE J.*, 38 (1992), 1969-1978.

[65] Kopaciewicz, W., Rounds, M.A., Fausnaugh, J. and Regnier, F.E., Retention model for high-performance ion-exchange chromatography, *J. Chromatogr.*, 266 (1983), 3-21.

[66] Rounds, M.A. and Regnier, F.E., Evaluation of a retention model for high-performance ion-exchange chromatography using two different displacing salts, *J. Chromatogr.*, 283 (1984), 37-45.

[67] Grager, R.R. and Regnier, F.E., Retention mechanism of lactate dehydrogenase in anion-exchange chromatography, *J. Chromatogr.*, 406 (1987), 237-246.

[68] Heng, M.H. and Glatz, C.E., Charged fusions for β-galactosidase retention in anion-exchange chromatography, *J. Chromatogr. A*, 689 (1995), 227-234.

[69] Chicz, R.M. and Regnier, F.E., Single amino acid contribution to protein retention in cation-exchange chromatography: Resolution of genetically engineered subtilisin variants, *Anal. Chem.*, 61 (1989), 2059-2066.

[70] Jandera, P. and Churacek, J., Gradient elution in liquid chromatography. I. The influence of the composition of the mobile phase on the capacity ratio (retention volume, band width, and resolution) in isocratic elution – theoretical considerations, *J. Chromatogr.*, 91 (1974), 207-221.

[71] Jandera, P. and Churacek, J., Gradient elution in liquid chromatography. II. Retention characteristics (retention volume, band width, resolution, plate number) in solvent – programmed chromatography – theoretical considerations, *J. Chromatogr.*, 91 (1974), 223-235.

[72] Ma, Z and Guiochon, G., Application of orthogonal collocation on finite elements in the simulation of non-linear chromatography, *Comput. Chem. Eng.*, 15 (1991), 415-426.

[73] Gallant, S. R., Vunnum, S. and Cramer, S. M., Optimization of preparative ion-exchange chromatography of proteins: linear gradient separations, *J. Chromatogr.*, 725 (1996), 295-314.

[74] Porath, J., Carlsson, J., Olsson, I. and Belfrage, G., Metal chelate affinity chromatography, a new approach to protein fractionation, *Nature*, 258 (1975), 598-599.

[75] Arnold, F.H., (editorial) *Methods*, 4 (1992), 1-3.

[76] Arnold, F.H., Metal-affinity separations: a new dimension in protein processing, *Bio/Technol.*, 9 (1991), 151-156.

[77] Sulkowski, E., Purification of proteins by IMAC, *Trends Biotechnol.*, 3 (1985), 1-7.

[78] Wong, J.W., Albright, R.L., and Wang, N.-H. L., Immobilized metal ion affinity chromatography (IMAC) – chemistry and bioseparation applications, *Sep. Purif. Methods*, 20 (1991), 49-106.

[79] Hochuli, E., Dobeli, H. and Schacher, A., New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues, *J. Chromatogr.*, 411 (1987), 177-184.

[80] Winserling, J.J., Berna, P., and Porath, J., How to use immobilized metal ion affinity chromatography, *Methods*, 4 (1992), 4-13.

[81] Zhao, Y.-J., Sulkowswi, E. and Porath, J., Surface topography of histidine residues in lysozymes, *Eur. J. Biochem.*, 202 (1991), 1115-1119.

[82] Arnold, F.H. and Haymore, B.L., Engineered metal-binding proteins: Purification to protein folding, *Science*, 252 (1991), 1796-1797.

[83] Sulkowski, E., The saga of IMAC and MIT, *Bioessays*, 10 (1989), 170-175.

[84] Josic, D., Reutter, W. and Reusch, J., Crown ethers as ligands for high-performance liquid chromatography of proteins and nucleic acids, *J. Chromatogr.*, 476 (1989), 309-318.

[85] Mantovaara, T., Pertoft, H. and Porath, J., Purification of human amyloid P component (SAP) by calcium affinity chromatography, *Biotechnol. Appl. Biochem.*, 11 (1989), 564-570.

[86] Porath, J., Olin, B., and Granstrand, B., Immobilized-metal affinity chromatography of serum proteins on gel-immobilized group III A metal ions, *Arch. Biochem. Biophys.* 225 (1983), 543-547.

[87] Andersson, L., Recognition of phosphate groups by immobilized aluminium(III) ions, *J. Chromatogr.*, 539 (1991), 327-334.

[88] Horstmann, B.J. and Chase, H.A., Modeling the affinity adsorption of immunoglobulin G to protein A immobilised to agarose matrices, *Chem. Eng. Res. Des.*, 67 (1989), 243-254.

[89] Patwardhan, A.V. and Ataai, M.M., IMA chromatography of a cell extract: Effect of loading pH on protein purification, *AIChE J.*, 43 (1997), 3232-3240.

[90] Mao, Q.M. and Hearn, M.T.W., Optimization of affinity and ion-exchange chromatographic processes for the purification of proteins, *Biotechnol. Bioeng.*, 52 (1996), 204-222.

[91] Jaroniec, M., Dabrowski, A. and Toth, J., Multilayer single-solute adsorption from dilute solutions on energetically heterogeneous solids, *Chem. Eng. Sci.*, 39 (1984), 65-70.

[92] Johnson, R.D. and Arnold, F.H., Multipoint binding and heterogeneity in immobilized metal affinity chromatography, *Biotechnol. Bioeng.*, 48 (1995), 437-443.

[93] Johnson, R.D. and Arnold, F.H., The Temkin isotherm describes heterogeneous protein adsorption, *Biochim. Biophys. Acta*, 1247 (1995), 292-297.

[94] Vunnum, S., Gallant, S.R., Kim, Y.J. and Cramer, S.M., Immobilized metal affinity chromatography: Modeling of nonlinear multicomponent equilibrium, *Chem. Eng. Sci.*, 50 (1995), 1785-1803.

[95] Vunnum, S. and Cramer, S., IMAC: Nonlinear elution chromatography of proteins, *Biotechnol. Bioeng.*, 54 (1997), 373-390.

[96] Ford, C.F., Suominen, I. and Glatz, C.E., Fusion tails for the recovery and purification of recombinant proteins, *Protein Expr. Purif.*, 2 (1991), 95-107.

[97] Lindner, P., Guth, B., Wulfing, C., Krebber, C., Steipe, B., Muller, F. and Pluckthun, Purification of native proteins from the cytoplasm and periplasm of *Escherichia coli* using IMAC and histidine tails: A comparison of proteins and protocols, *Methods*, 4 (1992), 41-56.

[98] Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. and Stuber, D., Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent, *Bio/Technol.*, 6 (1988), 1321-1325.

[99] Skerra, A., Pfitzinger, I. and Pluckthun, A., The functional expression of antibody F_v fragments in *Escherichia coli*: Improved vectors and a generally applicable purification technique, *Bio/Technol.*, 9 (1991), 273-278.

[100] Luther, J.R. and Glatz, C.E., Genetically engineered charge modifications to enhance protein separation in aqueous two-phase systems: Electrochemical partitioning, *Biotechnol. Bioeng.*, 44 (1994), 147-153.

[101] Luther, J.R. and Glatz, C.E., Genetically engineered charge modifications to enhance protein separation in aqueous two-phase systems: Charge directed partitioning, *Biotechnol. Bioeng.*, 46 (1995), 62-68.

[102] Zhao, J., Ford, C.F., Glatz, C.E., Rougvie, M.A. and Gendel, S.M., Polyelectrolyte precipitation of β-galactosidase fusions containing poly-aspartic acid tails, *J. Biotechnol.*, 14 (1990), 273-284.

[103] Hochuli, E. and Piesecki, S., Interaction of hexahistidine fusion proteins with nitrilotriacetic acid-chelated Ni^{2+} ions, *Methods*, 4 (1992), 68-72.

[104] Suominen, I., Ford., C., Stachon, D., Heimo, H., Niederauer, M., Nurmela, H. and Glatz, C., Enhanced recovery and purification of *Aspergillus* glucoamylase from *Saccharomyces cerevisiae* by the addition of poly(aspartic acid) tails, *Enzyme Microb. Technol.*, 15 (1993), 593-600.

[105] Dalboge, H., Dahl, H.-H.M., Pedersen, J., Hansen, J.W., and Christensen, T., A novel exzymatic method for production of authentic hGH from an Escherichia coli produced hGH-precursor, *Bio/Technol.*, 5 (1987), 161-164.

[106] Anba, J., Baty, D., Lloubes, R., Pages, J.-M., Joseph-Liauzun, E., Shire, D., Roskam, W. and Lazdunski, C., Expression vector promoting the synthesis and export of the human growth-hormone-releasing factor in *Escherichia coli*, *Gene*, 53 (1987), 219-226.

[107] Sharma, S.K., Evans, D.B., Vosters, A.F., Chattopadhyay, D., Hoogerheide, J.G. and Campbell, C.M., Immobilized metal affinity chromatography of bacterially expressed proteins engineered to contain an alternating-histidine domain, *Methods*, 4 (1992), 57-67.

[108] Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L. and Uhlen, M., Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in Staphylococci, *Gene*, 13 (1985), 1151-1162.

[109] Sassenfeld, H.M. and Brewer, S.J., A polypeptide fusion designed for the purification of recombinant proteins, *Bio/Technol.*, 2 (1984), 76-81.

[110] Brewer, S.J. and Sassenfeld, H.M., The purification of recombinant proteins using C-terminal polyarginine fusions, *Trends Biotechnol.*, 3 (1985), 119-122.

[111] Haymore, B.L., Bild, G.S., Salsgiver, W.J., Staten, N.R. and Krivi, G.G., Introducing strong metal-binding sites onto surfaces of proteins for facile and efficient metal-affinity purifications, *Methods*, 4 (1992), 25-40.

[112] Flaschel, E. and Friehs, K., Improvement of downstream processing of recombinant proteins by means of genetic engineering methods, *Biotechnol. Adv.*, 11 (1993), 31-78.

[113] Sassenfeld, H.M., Engineering proteins for purification, *Trends Biotechnol.*, 8 (1990), 88-93.

CHAPTER 2. PROCESS ENGINEERING STRATEGY FOR RECOMBINANT PROTEIN RECOVERY FROM CANOLA BY CATION EXCHANGE CHROMATOGRAPHY

A paper published in Biotechnology Progress

Chenming Zhang and Charles E. Glatz

Abstract

The suitability of canola as a recombinant protein production host was evaluated in terms of the potential for simple chromatographic recovery by ion exchange. To investigate the influence of the charge of a recombinant protein on recovery from canola, a series of mutants of T4 lysozyme of varying charge were spiked into canola protein extracts and this mixture was fractionated by cation-exchange chromatography. Two different means of charge modification were compared, namely point mutations and fusions.

Point mutations were proved the better means for adding positive charges. A linear relationship between the protein charge and the eluent conductivity, which could be used to guide the genetic engineering for protein recovery from canola, was found. It showed that an increase of +1 charge on T4 lysozyme increased the required conductivity (molarity) of eluent by 0.068 mS/cm (27.8 mM NaCl). For this specific case, T4 lysozyme with a nominal charge of +7 moves the point of elution into a valley between two major native canola protein peaks, where dramatic one-step enrichment can be obtained. Equivalent charge changes provided by polyarginine fusions gave very wide elution patterns that was ascribed

to either proteolytic degradation within the polyarginine fusion or to interaction of the polyarginine with polyanions present in canola.

While the above results came after a dialysis step to adjust the canola extract, elimination of the dialysis step did not significantly influence the purification behavior of the cation-exchange column. However, a more severe resin regeneration scheme was needed in order to maintain the column's performance.

Introduction

Plant agriculture is attracting attention for its potential to replace microorganisms for production of large quantities of recombinant proteins. Through genetic engineering of plants or "bio-farming", numerous recombinant proteins, have been expressed successfully (Whitelam, 1995; Goddijn and Pen, 1995; Ponstein et al., 1996). Because of the advantages of ease of scale up of production, established harvesting methods, and ease of storage and transport of the harvested plant material, bio-farming provides an economic way of producing biomass. However, the limited knowledge of expression control and lack of effective and economic downstream protein purification processes limits the use of plants as recombinant protein production hosts.

Canola, long used as an oil source and animal feed, has been explored as a possible host (van Rooijen and Moloney, 1995). The native canola proteins are mainly from two groups - a neutral, high molecular weight, 12S group and a basic, low molecular weight, 2S (or 1.7S) group (Bhatty et al., 1968; Schwenke, 1994). The 2S proteins have an isoelectric point of around 11 and have molecular weights between 12 to 17 kDa (Lönnerdal and Janson, 1972; Schwenke et al., 1973; Swanljung, 1970). The amino acid compositions of

proteins in this group are similar, and the proteins consist of two polypeptide chains linked by disulfide bridges (Lönnerdal and Janson, 1972). The main charge character comes from two exposed regions with high accumulation of the basic amino acids lysine and arginine (Schwenke, 1994).

Native canola proteins have been partly characterized by various chromatographic methods (Bhatty et al., 1968; Lönnerdal and Janson, 1972; Rabb et al., 1992; Amarowicz et al., 1993). With focus on the characterization of each individual native protein, a combination of different chromatographic methods, which most often included size exclusion and ion exchange chromatography, were employed in order to obtain the best possible resolution between chromatographic elution protein peaks. 2S proteins were shown to be chromatographically and electrophoretically homogeneous, and they were strongly adsorbed by carboxymethylcellulose (CM-cellulose) (Bhatty et al, 1968). Along with 2S proteins, there is another fraction of proteins, the intermediate fraction (IF), that can be weakly adsorbed by cation exchange at neutral pH (Rabb et al., 1992). These works showed that canola native proteins give relatively simple cation exchange chromatographic profiles. However, none of them evaluated canola seeds as a recombinant protein production host.

Chromatographic methods have been used in plant protein purification (Jervis and Pierpoint, 1989) because of their high resolving power, especially when very large production of a protein is not crucial. Ion-exchange chromatography (IEC) is widely used in purifying proteins of distinctive charges, and genetic engineering methods could be used to confer distinctive charges (Flachel and Friehs, 1993). This can be done by site-directed mutagenesis or attachment of a charged fusion tail. While the former method can best keep the integrity of a protein by avoiding unexpected proteolytic decomposition, the latter is more

flexible in "designing" the charged character without interfering in its stability or functioning (Ford et al., 1991; Wilkinson et al., 1995; Uhlen and Moks, 1990). In addition, the fusion method could be the only option when detailed structural information of a protein required for carrying out site-directed mutagenesis is not available.

Our goals here are to characterize the native canola proteins from the recombinant protein purification standpoint and investigate the effect of protein charge on a single step chromatographic purification. To investigate the charge effect, T4 lysozyme was chosen to be the target protein, because it and its mutants, both point mutation and fusions, are readily available in our lab. Cation exchange chromatography is adopted throughout this work because the target proteins are positively charged at neutral pH. The stability of different mutants in canola protein extract will also be addressed in this paper.

Materials and Methods

T4 lysozyme

T4 lysozyme is a basic protein with an isoelectric point above 9.0 and carries nine positive charges at neutral pH. It has 164 amino acids, and its molecular weight is 16,800 kDa. The T4 lysozyme mutants used in this study were developed using both site-directed mutagenesis and fusions (Table 2.1). The point mutants were developed by replacing surface lysine residues with glutamic acids (Sun et al., 1991). The three mutants are K16E (single mutant), K16/135E (double mutant), and K16/135/147E (triple mutant), respectively, with the number designating the position of the replaced lysines to give variants with charges of +7, +5, and +3 at neutral pH. The fusions were polyarginine tails of 2 (U1) or 4 (U2)

Table 2.1. Charge of T4 lysozyme mutants and fusions at pH=7.0

T4 lysozyme	Net charge
Wild type	9.0
K16E, single mutant	7.0
K16/135E, double mutant	5.0
K16/135/147E, triple mutant	3.0
U1, 2 arginines	5.0
U2, 4 arginines	7.0

arginines attached to the C-terminal of the triple mutant (Fan and Glatz, 1998) providing net charges of +5 and +7, respectively.

Enzyme production and protein assays

T4 lysozyme production was induced in *Escherichia coli* by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). Extracted protein was purified by cation exchange chromatography (Luther and Glatz, 1994).

Lysozyme activity was measured by the clearing of a *Micrococcus lysodeiktikus* cell suspension (Perry et al., 1965). The concentration of protein was obtained by using Bio-Rad protein assay (Bio-Rad, Hercules, CA); the pure T4 lysozyme solution content was also determined by A_{280} based on $e_{280}^{1\%}$ =1.28 (Poteete et al., 1991).

Preparation of protein extract

Non-transgenic canola seed was provided by Pioneer Hi-bred International (Johnston, IA). Canola seed was ground, dehulled, and flaked before oil extraction. The defatted

canola meal was obtained by two stages, 30 minutes each, of cold *n*-hexane extraction with a ratio of 1g seed/20 ml solvent. The final oil content of the canola meal is < 2% (w/w).

Native canola proteins were extracted into 50 mM sodium phosphate (NaPi) buffer (pH=7.0) at 1:10 (w/v) for 30 minutes with mild stirring. The protein concentration of the extract after centrifugation and filtration by a 0.22 μm syringe filter was 30 ± 5 mg/ml. Then the protein extract was dialyzed (1:400; Spectra/Pro®1 membrane tube; MWCO 6-8,000; Spectrum, Laguna Hills, CA) at 4°C for at least 36 hours against 20 mM NaPi (pH=7.0), which was the equilibrating buffer for ion exchange chromatography, with one fresh buffer change at around 18 hours into dialysis. After dialysis, the extract was filtered through a 0.22 μm syringe filter again to remove precipitates developed during dialysis. The protein concentration of the final filtrate was 12 ± 1 mg/ml.

Chromatography

The flow chart of the experimental procedures is shown in Figure 2.1. Native canola proteins were characterized by both cation-exchange and anion-exchange chromatography. Cation-exchange chromatography (CIEC) was then used throughout since the target protein, T4 lysozyme, is positively charged at the experimental condition of pH 7.0. During each chromatographic run, 6 ml of sample was loaded to the column. After 60 ml of washing by 20 mM NaPi, pH 7.0 (buffer A), linear gradient elution from buffer A to 20 mM NaPi with 500 mM NaCl, pH=7.0 (buffer B) was applied in order to elute bound proteins. Meanwhile 5 ml fractions were collected for later activity assay to identify where the spiked T4 lysozyme was eluted. 1 ml/min flow rate was used for all the chromatographic experiments.

As a simpler alternative, dialysis was eliminated and the canola protein extract was loaded to a column directly after centrifugation and filtration (Fig. 2.1). All other parameters were unchanged.

Column regeneration was accomplished by serial washing with 20 ml of buffer B, 20 ml 0.1M HCl, 20 ml of buffer B (20/20/20), and then washing to baseline by buffer A. After using non-dialyzed sample, regeneration volumes were increased to 40/40/40, respectively.

All columns were packed to a bed height of 6.5± 0.2 cm in 1x10 Econo columns (Bio-Rad, Hercules, CA). Chromatography was performed by using a Fast Performance Liquid Chromatography (FPLC) system controlled by BioLogic software (Bio-Rad, Hercules, CA).

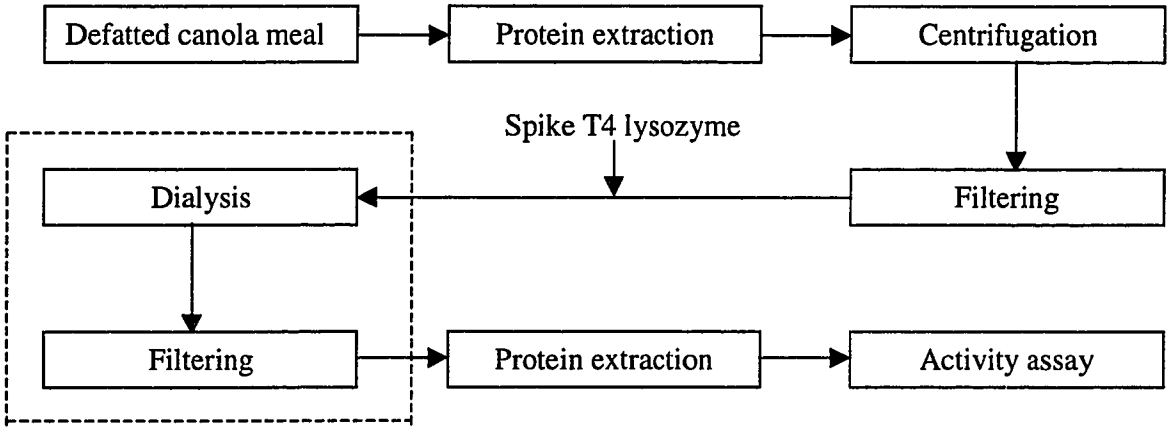


Figure 2.1. Flow chart for the extraction and fractionation. One alternative eliminates the dialysis step (the dashed line box)

Chemicals and materials

All chemicals were purchased from Fisher (Itasca, IL). *Micrococcus lysodeiktikus* and all the resins were purchased from Sigma (St. Louis, MO).

Results and Discussion

Characterization of native canola proteins

Figure 2.2 shows the chromatographic profiles of native canola protein for cation and anion-exchange chromatography. Both chromatograms show rather simple elution patterns of native canola proteins. With cation-exchange chromatography, all bound canola proteins can be eluted into two major peaks with a wide, low background valley in between. This provides a potential elution site for a genetically engineered protein. With anion-exchange chromatography, there are three visible native canola protein peaks. Gaps between the peaks also provide potential elution sites for negatively charged proteins at neutral pH. Moreover, both of the chromatograms show patterns similar to those obtained by Finlayson (Finlayson, 1966) who used a pH gradient in both weak anion and cation exchange chromatography.

From our experimental results (Fig. 2.3), the only difference between using strong or weak cation exchange chromatography is the ionic strength at which the peaks were eluted, even though CM type resin does provides a broader valley to target a protein. The simple chromatographic elution behavior of 2S proteins of canola (Bhatty et al., 1968) is evident with a single elution peak eluted at high salt concentration.

T4 lysozyme purification

Figure 2.4 shows the elution profiles of canola samples spiked with different point-mutated T4 lysozymes, and Figure 2.5 shows the corresponding activity assay results.

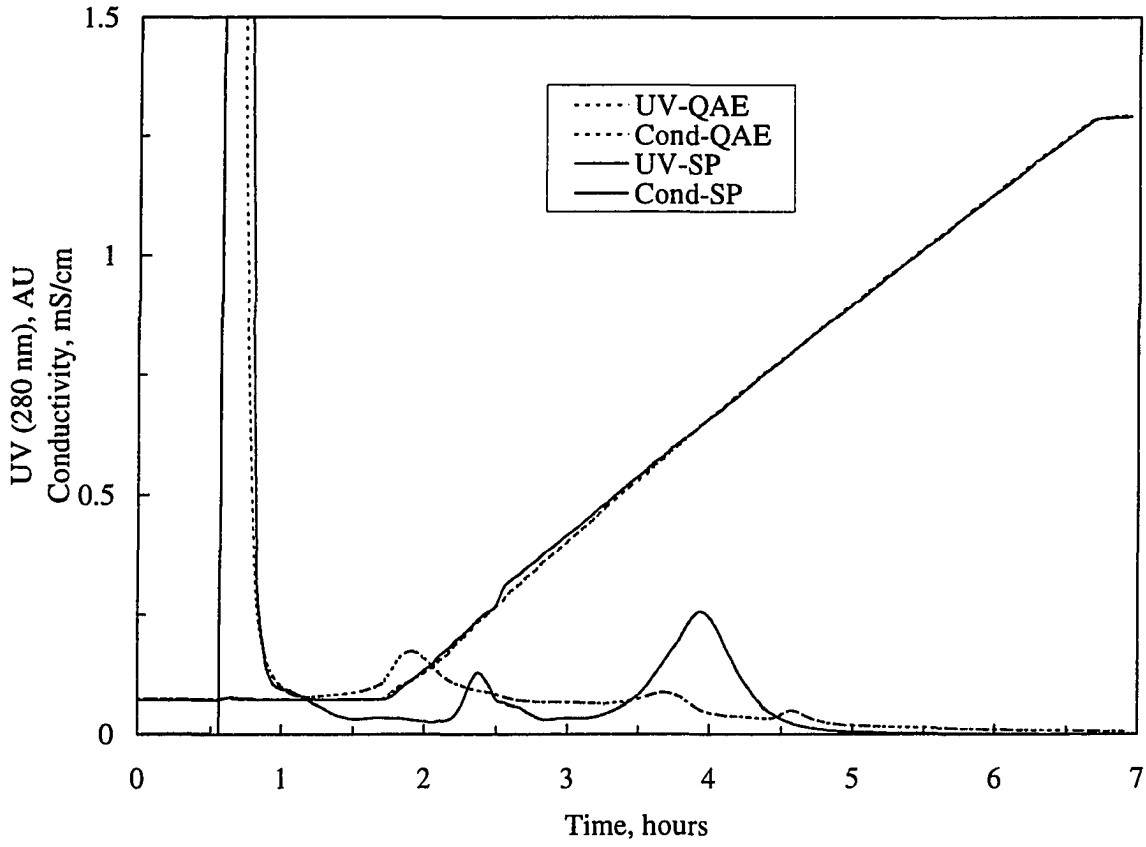


Figure 2.2. Comparison of native canola protein elution profiles between cation and anion-exchange chromatography. Cation exchanger: Sulfopropyl Sephadex (SP); anion exchanger: Diethyl[2-hydroxypropyl]aminoethyl Sephadex (QAE). All the chromatographic experiments were carried out using same linear gradient program.

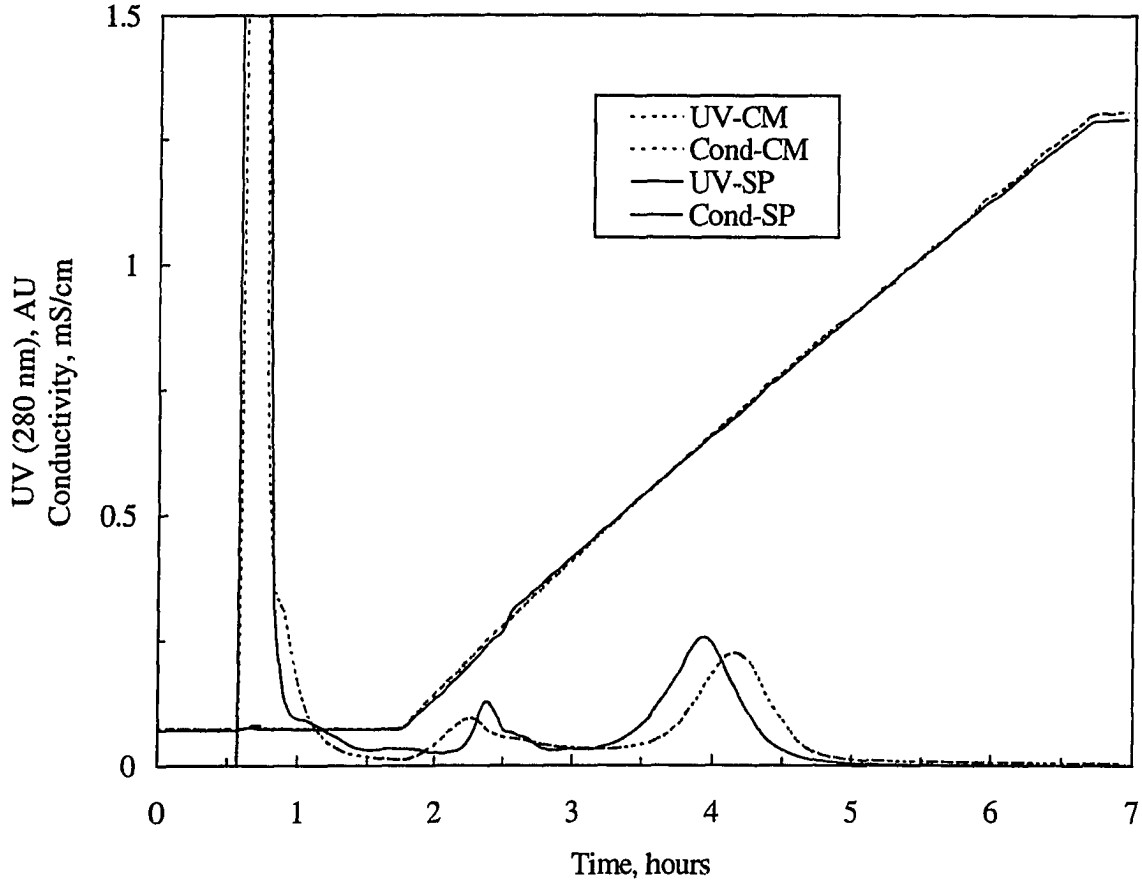


Figure 2.3. Comparison of weak (CM) and strong (SP) cation exchange columns in separation of native canola proteins. CM = Carboxymethyl Sephadex.

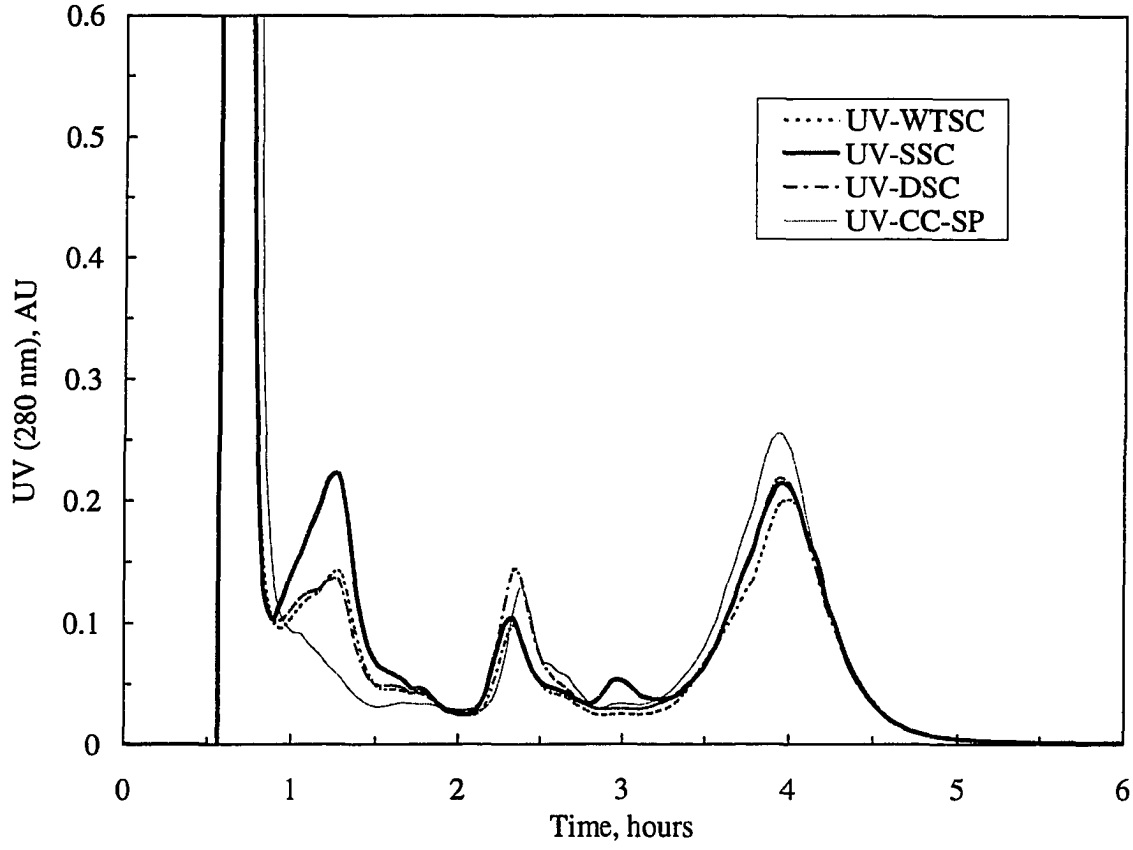


Figure 2.4. Comparison of chromatograms of different canola samples with spiked T4 lysozymes by cation exchange chromatography. 2 ml of T4 lysozyme stock solution (~1mg/ml) was added into 8 ml of canola protein extract. Cation exchanger: SP. CC: Canola Control; WTSC: Wild-Type T4 lysozyme Spiked Canola sample; SSC: Single mutant T4 lysozyme (K16E) Spiked Canola sample; DSC: Double mutant T4 lysozyme (K16/135E) Spiked Canola sample. Linear gradient started at 1.7 hours. The peak before the gradient started was due to incomplete dialysis. Repeated experiment with 72 hours of dialysis showed no such a peak.

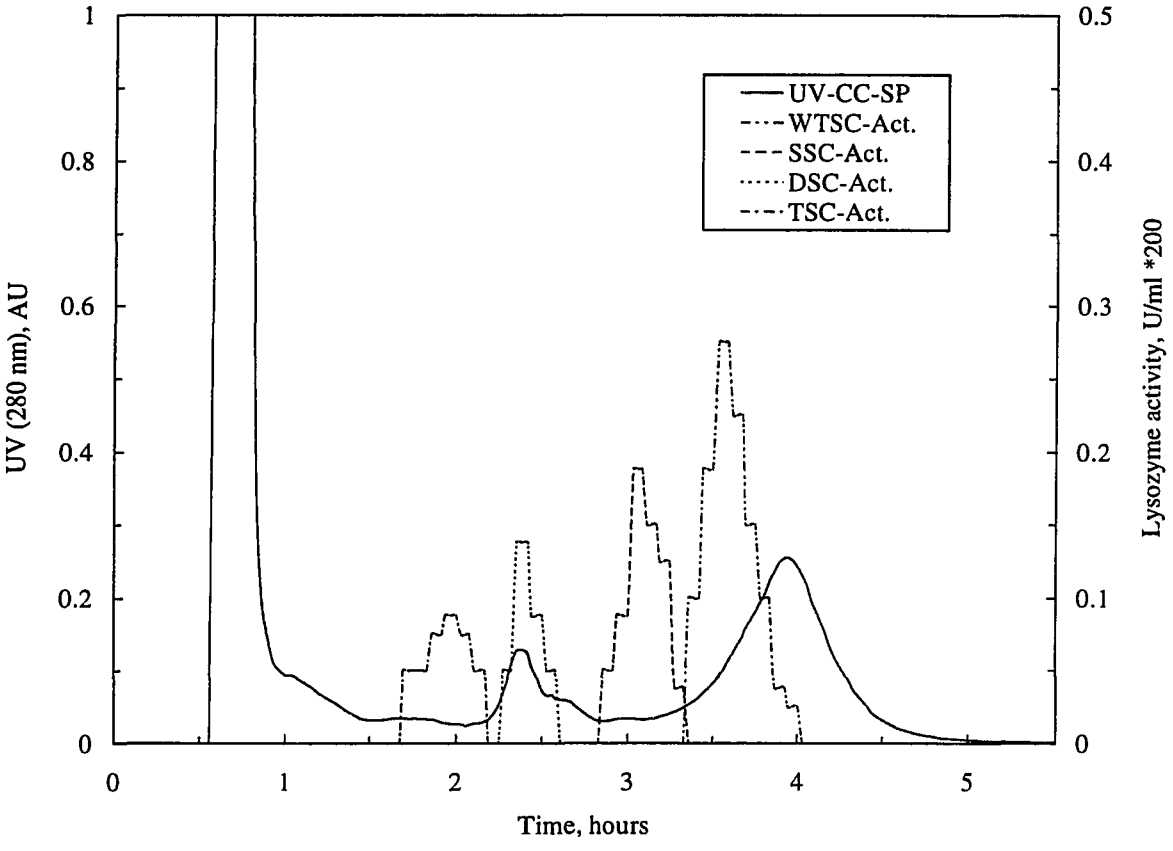


Figure 2.5. Superposition on the native canola profile of lysozyme activity assays from separate injections of each of the point mutant series spiked canola extract. Absorbance at 280 nm of canola extract (UV-CC) is shown at the left ordinate, and the activity (Act.) of lysozyme is shown at the right ordinate. TSC: Triple mutant T4 lysozyme (K16/135/147E) Spiked Canola sample. Other nomenclatures see Fig. 2.4 caption.

From Fig. 2.4 and 2.5, charges of T4 lysozymes have a significant influence on their retention behavior. Single mutant T4 lysozyme spiked canola sample gives a noticeable peak in the space between the two native canola protein peaks (Fig. 2.4), while other point mutants overlap with one or the other of the two native peaks. 60% of the single mutant T4 lysozyme loaded to the column was recovered from this peak, which also had 90% of the lysozyme activity detected in fractions collected after gradient elution started. Fig. 2.6 shows the gel electrophoresis result of the IEC fractions of single mutant T4 lysozyme spiked canola sample. Besides the lysozyme band, there are four detectable bands indicating four contaminating proteins. Two of them with a molecular weight around 35 kDa (too weak to be seen) and the other two with a molecular weight less than 14 kDa (one of them can be seen in Fig. 2.6, and the other one has a slightly higher molecular weight). However, the estimated lysozyme purity from the band intensity on the gel was higher than 90%.

On the other hand, the fusion-spiked canola extracts show lysozyme activity spread over many fractions with two main activity peaks (Fig. 2.7). U2 T4 lysozyme, which carries the same charge as single-mutant T4 lysozyme, does show an activity peak at the same location where single-mutant T4 lysozyme would elute. However, only 66% of total lysozyme activity was recovered in this peak compared to 90% for single-mutant T4 lysozyme activity. This indicates a low recovery for the fusion strategy. Of the two main activity peaks seen for each fusion protein (Fig. 2.7), the later one corresponds to where the point mutant with the same charge would be eluted, while the earlier corresponds to that of triple mutant T4 lysozyme without the fusion tail. Evidently, the tails were degraded by proteases in the canola extract. A less likely explanation is that polyanions found in canola, mainly pectic acid and phytic acid (Schwenke et al., 1991; Evans et al., 1982; Siy and Talbot,

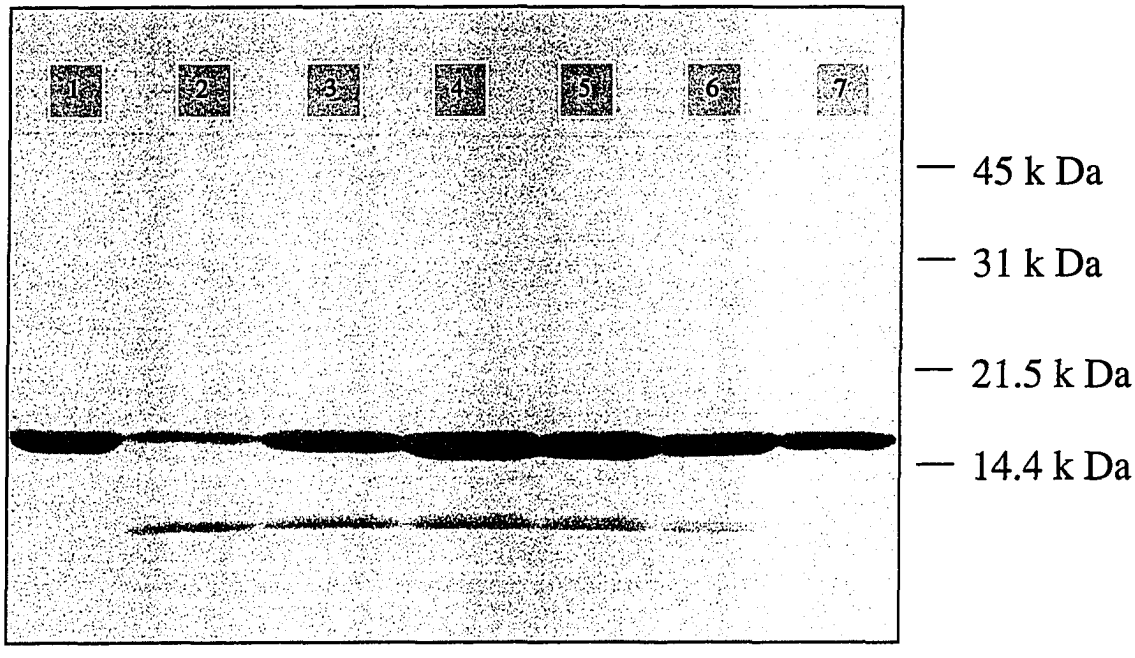


Figure 2.6. SDS-PAGE of those fractions from IEC of canola extract spiked with single mutant T4 lysozyme (K16E) showing activity. Lane 1: Single mutant T4 lysozyme stock solution; Lane 2-7: IEC fractions comprising the peak in Fig. 2.4 (UV-SSC). Lane 4 corresponds to the activity peak in Fig. 2.5 (SSC-Act.). All fractions were concentrated about 26-fold before loading to the gel.

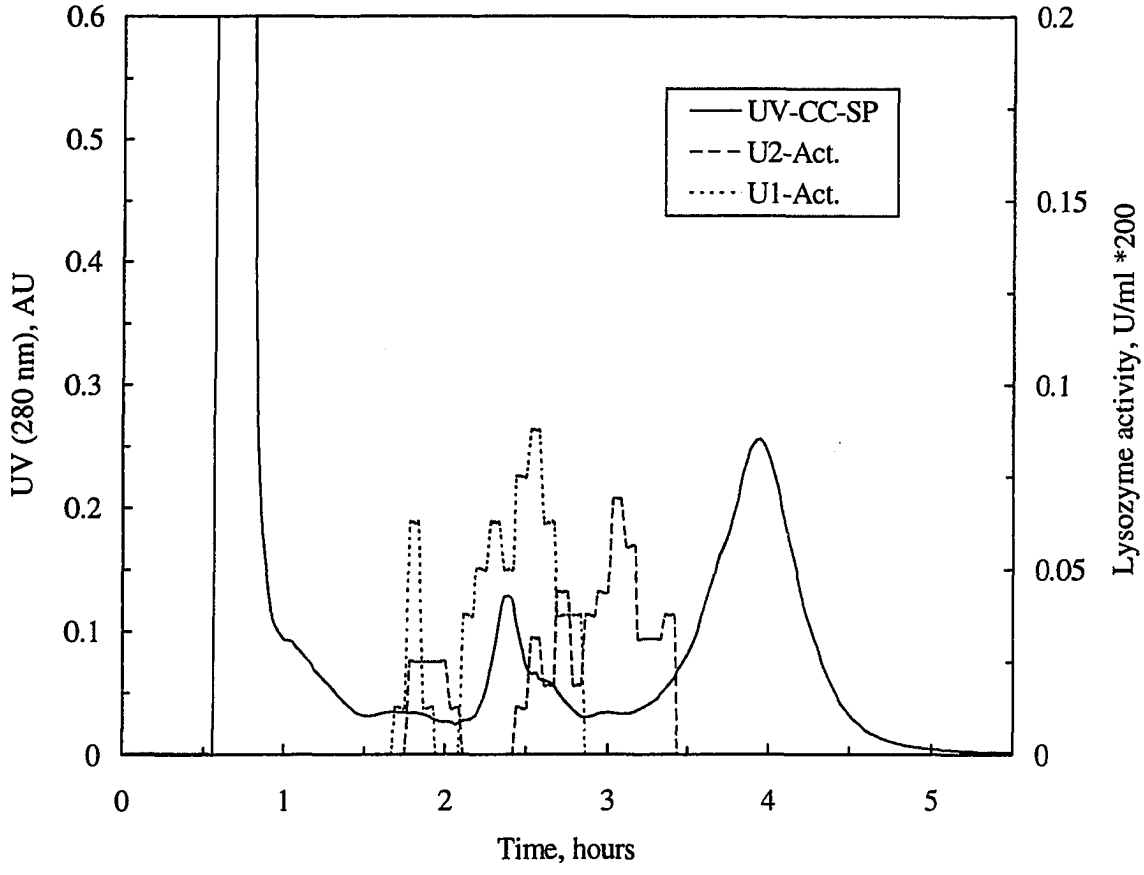


Figure 2.7. Relative retention times of fusion T4 lysozymes (U1 and U2) compared to that of native canola proteins. Cation exchanger: SP.

1982), are present in the extract and interfere with binding the positively charged fusion proteins and the ion exchanger. Pectic acid (cellulose sulphates) and phytic acid (poly phosphate) binding to the high charge density area of the tail would block the intended stronger binding of fusions to the column.

Nevertheless, T4 lysozyme and its mutants demonstrate the ability to improve the ease of purification from canola by manipulation of charge. Single-mutant T4 lysozyme with a positive charge of 7 at neutral pH can be enriched 35 times from a canola protein extract by this single chromatographic step. Other point-mutant T4 lysozymes were eluted with one of the two major native canola protein peaks. Subjected to the influence of polyanions and protease activity, polyarginine fusions do not show advantages over point mutations in the canola system. An alternative would be to use polyanionic fusions, which in microbial systems have not had the same susceptibility to proteolysis as polyarginine fusions (Ford et al., 1991; Zhao et al., 1990).

Figure 2.8 shows the relationship between the conductivity at elution (y, mS/cm) and the charge at pH 7 (x) of each mutant. The results are fitted by linear relationship on point mutant and WT T4 lysozymes,

$$y = -0.078 + 0.068 * x \qquad (R^2 = 0.99)$$

Based on the above equation, one additional positive charge on a protein would increase the conductivity of an eluent by 0.068 mS/cm, which corresponds to an increase of 27.8 mM NaCl in 20 mM NaPi (27.8 mM NaCl/charge). Previous studies on human urogastrone (Sassenfeld and Brewer, 1984) and β–galactosidase (Zhao et al., 1990; Heng and Glatz, 1993) showed different increments in elution buffer ionic strength per added charge. In the

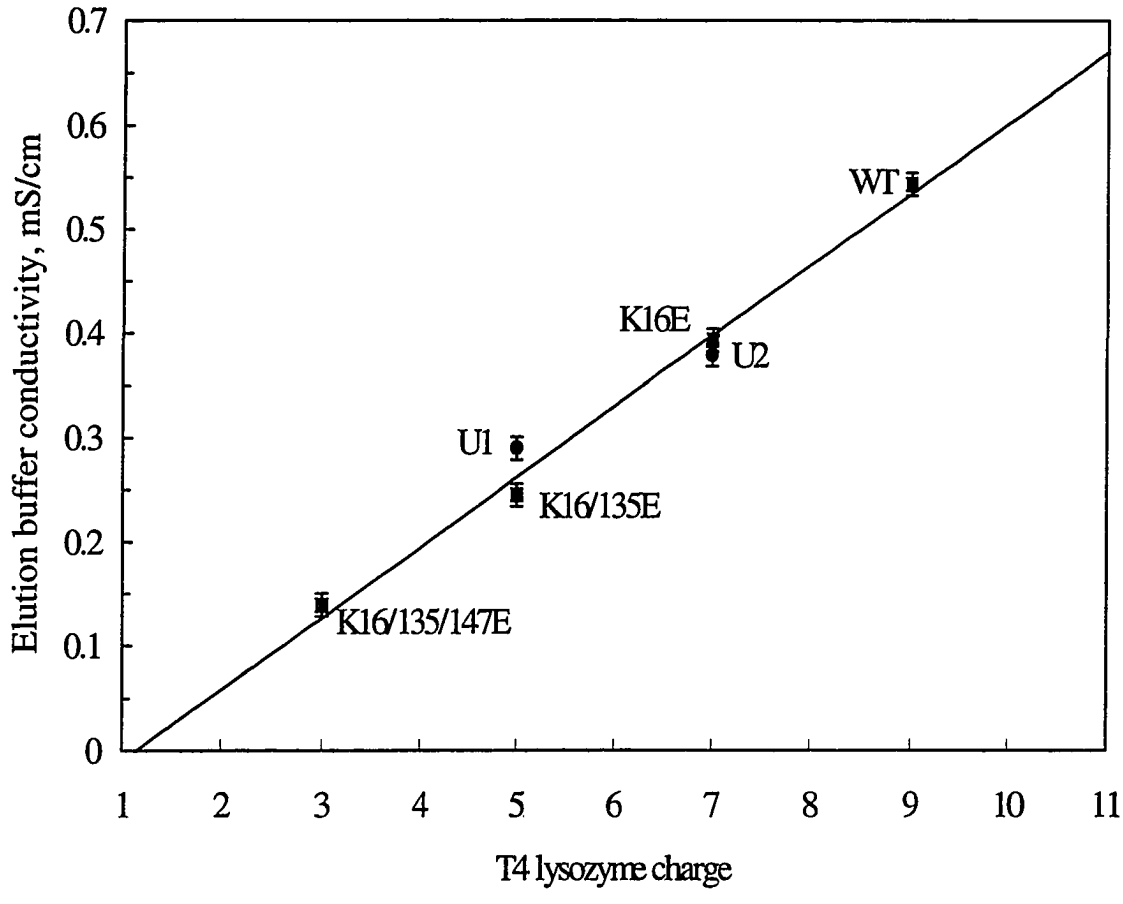


Figure 2.8. Eluent conductivity at which each T4 lysozyme is eluted at pH 7.0, and the linear regression is performed among values obtained from point mutant series.

case of human urogastrone, five additional arginines at the carboxyl terminus increased the concentration of NaCl of eluent from 50 mM to 400 mM in Tris/urea buffer (5M urea, 40 mM Tris-acetate, pH 5.5), which corresponds to 70 mM/charge. In the case of β–galactosidase, polyaspartic acid tails with 1, 5, and 11 of aspartic acid residues at the carboxyl terminus provided additional charges to change the elution profile of β–galactosidase in anion-exchange chromatography (AEC) (Zhao et al., 1990) and hollow fiber ion-exchange membrane adsorption (HFIEM) (Heng and Glatz, 1993). In AIEC, an average of 18.2 mM NaCl/charge (0.1 M NaPi, pH 6.0) of increase was reported, and for HFIEM, the increase per additional charge is 13.6 mM NaCl (0.089 M NaPi, pH 5.7) during elution. These are summarized in Table 2.2. The difference in increment of NaCl per additional charge during elution is probably due to the pH of buffers used, systems which proteins are to be purified from, the protein size, and the retention medium used in separation.

Nonetheless, with the above equation, a reasonable prediction of when a protein would be eluted from the column could be made given the charge it carries. This can guide how the genetic engineering work should be done for a recombinant protein which would be produced with canola as the production host. There are two targeting elution sites for positively charged recombinant proteins; one is between the two native canola protein peaks, and the other is after the second native canola peak. If without altering the identity (secondary structure and activity) of a protein, the protein can be genetically engineered to carry +13 charge, from the relationship we obtained, it would be eluted at a buffer conductivity of 0.81±0.01 mS/cm, which corresponds to a salt concentration of 300±4 mM NaCl in 20 mM NaPi. When this charge level is not possible, the protein can be targeted to the gap between the native canola protein peaks with a charge of +7. However, since this

Table 2.2. Comparison of the increment in salt concentration per added charge for protein elution.

Protein	Human Urogastrone	β–galactosidase		T4 lysozyme
Molecular weight (k Da)	6	116.4		16.8
Separation method	CEC	AEC	HFIEM	CEC
System	*E. coli* fermentation	*E. coli* fermentation	*E. coli* fermentation	Canola extract
Buffer and pH	5 M Urea, 40 mM Tris-acetate, pH 5.5	0.1 M NaPi, pH 6.0	0.089 M NaPi pH 5.7	0.02 M NaPi pH 7.0
Elution method	Gradient	Step	Step	Gradient
Δ mM NaCl/Charge	70	18.2	13.6	27.8
Reference	Sassenfeld and Brewer, 1984; Gregiry and Preston, 1977	Zhao et al., 1990	Heng and Glatz, 1993	This work

equation was obtained by studying a relatively small protein (~17 kDa), the application to large proteins should be done relative to the observed elution point of the unaltered protein.

Process modification

Eliminating the dialysis step for buffer exchange would simplify the process and shorten the processing time. Two alternatives were examined to eliminate the dialysis step: one was to change the protein extraction buffer to the column equilibrating buffer, and the other was to eliminate the dialysis without changing the extraction buffer.

Figure 2.9 is the comparison between the chromatograms of canola protein samples extracted by 20 mM NaPi and 50 mM NaPi (pH=7.0), respectively, but without dialysis before loading (Fig. 2.1). The total protein extracted at 20 mM NaPi is 25% less than that at 50 mM NaPi, but from the peak areas in the chromatograms in Fig. 2.9 (peak 1 and 2), the total basic proteins extracted are similar. Nevertheless, the space between the two native canola peaks remains the same for both extraction buffers used.

Figure 2.10 is the chromatogram of the single-mutant T4 lysozyme-spiked canola protein extract (50 mM NaPi) without the dialysis step. The spiked single-mutant T4 lysozyme still elutes between the two native canola protein peaks. The single-mutant peak shifts to the right in Fig. 2.10 compared to Fig. 2.3, because the washing time after loading the sample was increased by 120 minutes (1ml/min) in order to obtain a good baseline before the linear gradient was started.

Compared to dialyzed samples (Fig. 2.2), the non-dialyzed canola samples (Fig. 2.9) show a larger peak after loading the samples and before the gradient elution was started. This group of proteins was partially eliminated during sample dialysis largely due to precipitation, and the precipitation was significant at the end of dialysis. Gel electrophoresis

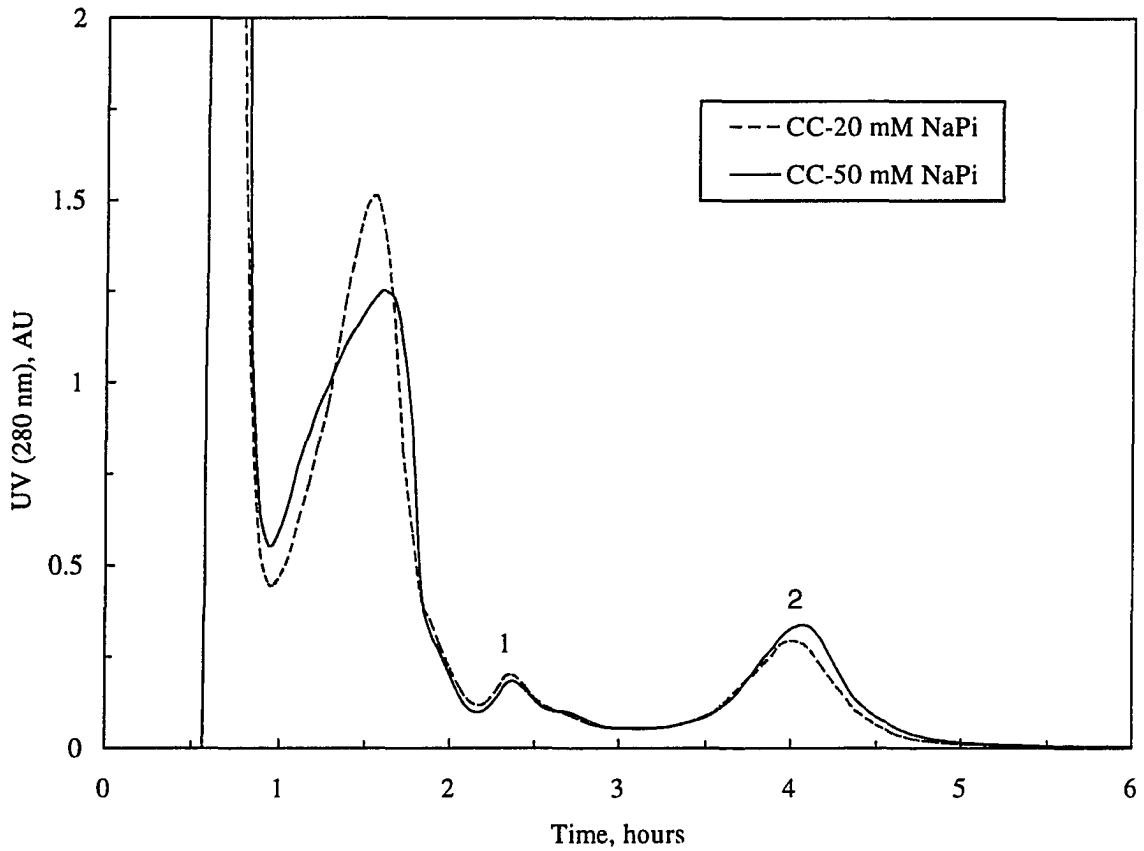


Figure 2.9. Comparison of different extraction buffers when dialysis was not performed before IEC. Cation exchanger: SP. Linear gradient elution started at ~1.7 hour.

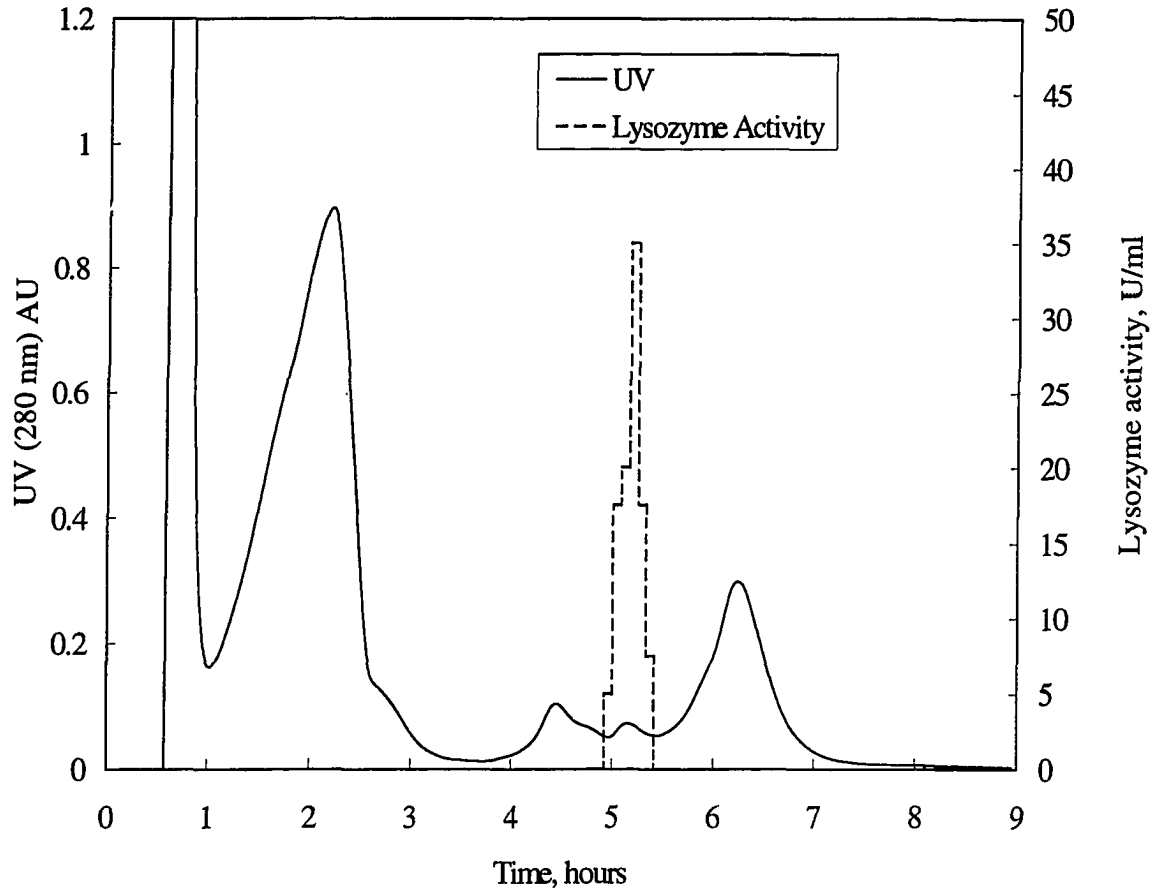


Figure 2.10. IEC profile of single mutant T4 lysozyme spiked canola sample with 50 mM NaPi as the extraction buffer and without dialysis before IEC. Cation exchanger: SP. Sample loaded to column was 6 ml. Washing time after loading was 180 minutes. Linear gradient started at ~3.5 hour.

(not shown here) on fractions with T4 lysozyme activity (Fig. 2.10) shows higher intensities than the corresponding fractions of dialyzed sample (UV-SSC in Fig. 2.4, SSC-Act. in Fig. 2.5) on corresponding bands in Fig. 2.6. In addition, there were four more very weak bands for the fractions of non-dialyzed sample, two at 30 kDa and the other two around 66 kDa. So, the dialysis step not only exchanges the buffer to a lower ionic strength (50 mM NaPi to 20 mM NaPi) but also acts as a pre-separation step to eliminate some native canola proteins, predominantly some weakly binding proteins. However, the eliminated proteins do not influence the purification performance of a highly charged (such as +7.0 or more) recombinant protein except slightly higher concentrations of contaminating proteins in corresponding fractions. In addition, the recovery of single-mutant T4 lysozyme from a non-dialyzed sample (>80%) is significantly higher than the recovery from a dialyzed sample (~60%). This is probably because some T4 lysozyme was precipitated during dialysis.

The drawback of no dialysis is that the fouling of the column was severe, which was mainly due to plugging by various precipitates, e.g. lipids, proteins, and other fine particles. This altered the flow rate/pressure drop through the column. Furthermore, when 50 mM NaPi buffer was used in extraction, the column fouling was more serious, since the equilibrating buffer was 20 mM NaPi. The decrease of the salt concentration might cause some proteins to precipitate once the sample was loaded to the column. Regeneration with 0.5 M NaOH was needed to recover column performance.

Conclusions

Canola as a recombinant protein production host offers opportunity for selective recovery of the recombinant protein by IEC. A charge change of -2 in wild-type T4

lysozyme permits its selective recovery in a single cation exchange chromatographic step. A linear relationship between the protein charge and the eluent ionic strength was found. This could be used to guide a genetic engineering work on a recombinant protein with canola as a host. Charge distribution on T4 lysozymes is not very important from our results. Fusions and point mutations have the same retentions when their charges are the same. However, performance of the polyarginine fusions is severely compromised by canola extract components.

Dialysis of extracts acted as a pre-purification step by eliminating some proteins weakly bound to the cation exchanger as a result of precipitation during the dialysis step. However, the dialysis step could be eliminated in order to simplify and shorten the purification processes. The same separation performance was still achieved in the case of purifying single-mutant T4 lysozyme, but column fouling increased.

Notation

AIEC	Anion exchange chromatography
CC	Canola control
CIEC	Cation exchange chromatography
CM	Carboxymethyl Sephadex, weak cation exchanger
DSC	Double mutant T4 lysozyme (K16/135E) spiked canola sample
HFIEM	Hollow fiber ion-exchange membrane adsorption
QAE	Diethyl[2-hydroxypropyl]aminoethyl Sephadex, strong anion exchanger
SP	Sulfopropyl Sephadex, strong cation exchanger
SSC	Single mutant T4 lysozyme (K16E) spiked canola sample

TSC	Triple mutant T4 lysozyme (K16/135/147E) spiked canola sample
U1	Triple mutant T4 lysozyme with a fusion tail of 2 arginines
U2	Triple mutant T4 lysozyme with a fusion tail of 4 arginines
WTSC	Wild-Type T4 lysozyme spiked canola sample

Acknowledgment

This work is supported by NSF under Grant No. BES9522644. We thank Pioneer Hi-Bred International, Inc., Johnston, IA, for providing canola seed. We also thank Dr. Zivko Nikolov and Dr. Ann Kusnadi for their suggestions and help.

Literature Cited

Amarowicz, R.; Ciska, E.; Kmita-Glaewska, H. Fast ionexchange chromatography of low molecular rapeseed proteins. *Pol. J. Food Nutr. Sci.* **1993**, *2/43*, 77-82.

Bhatty, R.S.; McKenzie, S.L.; Finlayson, A.J. The proteins of rapeseed (Brassica napus L.) soluble in salt solution. *Can. J. Biochem.* **1968**, *46*, 1191-1197.

Evans, W.J.; McCourtney, E.J.; Shrager, R.I. Titration studies of phytic acid. *J. Am. Oil Chemists Soc.* **1982**, *59*, 189-191.

Fan, W.; Glatz, C.E. Contribution of protein charge to partition in aqueous two-phase systems. *Biotechnol. Bioeng.* **1998**, in press.

Finlayson, A.J. Incorporation of ^{14}C into two proteins from rapeseed (Brassica Napus L.). *Can. J. Biochem.* **1966**, *44*, 997-1004.

Flaschel, E.; Friehs, K. Improvement of downstream processing of recombinant proteins by means of genetic engineering methods. *Biotech. Adv.* **1993**, *11*, 31-78.

Ford, C.; Suominen, I.; Glatz, C. E. Fusion tails for the recovery and purification of recombinant proteins. *Prot. Expr. Purif.* **1991**, *2*, 95-107.

Goddijn, O.J.M.; Pen, J. Plants as bioreactors. *Trends Biotechnol.* **1995**, *13*, 379-387.

Gregiry, H.; Preston, B.M. The primary structure of human urogastrone. *Int. J. Peptide Protein Res.* **1977**, *9*, 107-118.

Heng, M.H.; Glatz, C.E. Charged fusions for selective recovery of β–galactosidase from cell extract using hollow fiber ion-exchange membrane adsorption. *Biotechnol. Bioeng.* **1993**, *42*, 333-338.

Jervis, L.; Pierpoint, W.S. Purification technologies for plant proteins. *J. Biotech.* **1989**, *11*, 161-198.

Lönnerdal B.; Janson, J.-C. Studies on Brassica seed proteins. *Biochim. Biophys Acta* **1972**, *278*, 175-183.

Luther, J.R.; Glatz, C.E. Genetically engineered charge modifications to enhance protein separation in aqueous two-phase systems: Electrochemical partitioning. *Biotechnol. Bioeng.* **1994**, *44*, 147-153.

Perry, R.M. Jr.; Chandan, R.C.; Shahani, K.M. A rapid and sensitive assay of muramidase. *Proc. Soc. Exp. Biol. Med.* **1965**, *119*, 384-6.

Ponstein, A.S.; Verwoerd, T.C.; Pen, J. Production of Enzymes for industrial use. *Ann. NY Acad. Sci.* **1996**, *792*, 91-98.

Poteete, A.R.; Sun, D.-P.; Nicholson, H.; Mattews, B.W. Second-Site revertants of an inactive T4 lysozyme mutant restore activity by restructuring the active left. *Biochemistry* **1991**, *30*, 1425-1432.

Raab, B.; Leman, H.; Schwenke, K.D.; Kozlowaka, H. Comparative study of the protein patterns of some rapeseed (Brassica napus L) varieties by means of polyacrylamide gel electrophoresis and high-performance liquid chromatography. *Nahrung* **1992**, *36*, 239-247.

Sassenfeld, H.M.; Brewer, S.J. A polypeptide fusion designed for the purification of recombinant proteins. *Bio/Technol.* **1984**, *2*, 76-81.

Schwenke, K.D. In *New and developing sources of food proteins*; Hudson, B.J.F., Ed.; Chapmam & Hall: New York, 1994; pp 281-306.

Schwenke, K.D.; Mothes, R.; Gast, K.; Zirwer, D. Rapeseed protein - polyanion interactions. Protein Interaction. Symp. 201st Annu. Meet. Am. Chem. Soc. **1991**, 253-267.

Schwenke, K.D.; Raab, B.; Uhlig, J.; Böttger M.; Freimuth U. Seed proteins. 3. The isolation and characterization of albumins from sunflower seed and rapeseed. *Nahrung* **1973**, *17*, 791-809.

Siy, R.D.; Talbot, F.D.F. Preparation of low-phytate rapeseed protein by ultrafiltration: I. The aqueous extraction of phytate from deoiled rapeseed meals. *J. Am. Oil Chemists Soc.* **1982**, *59*, 191-194.

Sun, D.-P.; Söderlind, E.; Basse, W.A.; Wozniak, J.A.; Sauer, U.; Mattews, B.W. Cumulative site-direct charge-change replacements in bacteriophage T4 lysozyme suggest that long-range electrostatic interactions contribute little to protein stability. *J. Mol. Biol.* **1991**, *221*, 873-887.

Swanljung, P. Purification and some properties of a group of small basic proteins from rapeseed. *Acta Chem. Scand.* **1970**, *26*, 405-407.

Uhlen, M.; Moks, T. Gene fusions for purpose of expression: An introduction. *Methods Enzymol.* **1990**, *185*, 129-143.

Van Rooijen, G.J.H.; Moloney, M.M. Plant seed oil-bodies as carriers for foreign proteins. *Bio/Technology* **1995**, *13*, 72-77.

Whitelam, G.C. The production of recombinant proteins in plants. *J. Sci. Food Agric.* **1995**, *68*, 1-9.

Wilkinson, D.L.; Ma, N.-T.; Haught, C.; Harrison, R.G. Purification by immobilized metal affinity chromatography of human atrial natriuretic peptide expressed in a novel thioredoxin fusion protein. *Biotechnol. Prog.* **1995**, *11*, 265-269.

Zhao, J.; Ford, C.F.; Glatz, C.E.; Rougvie, M.A.; Gendel, S.M. Polyelectrolyte precipitation of β–galactosidase fusions containing poly-aspartic acid tail. *J. Biotechnol.* **1990**, *14*, 273-284.

CHAPTER 3. GENETIC ENGINEERING STRATEGIES FOR PURIFICATION OF RECOMBINANT PROTEINS FROM CANOLA BY ANION EXCHANGE CHROMATOGRAPHY: AN EXAMPLE OF β-GLUCURONIDASE

A paper submitted to Biotechnology Progress

Chenming Zhang[1], Robert T. Love[2], Joseph M. Jilka[2], and Charles E. Glatz[3]

Abstract

The elution behavior of native canola proteins from different anion exchange resins was determined. The elution profiles showed the potential for simplified recovery of acidic recombinant proteins from canola. When Q-sepharose fast flow was used, there were three possible sites to target the elution of a recombinant protein for minimal contamination with native proteins. The feasibility of targeting these sites was assessed using the acidic protein β-glucuronidase (GUS/GUSD0) along with three polyaspartate fusions to the wild-type *Escherichia coli* GUS. The fusions contained 5 (GUSD5), 10 (GUSD10), or 15 (GUSD15) aspartic acids fused to the C-terminus and were chosen to extend the elution time. The three fusions and the wild-type were produced in *E. coli*, purified, and added to canola extracts before chromatography. As a control to assess the equivalence of behavior of a protein produced in transgenic canola and these spiked experiments, an extract of transgenic canola

---

[1] Graduate student, primary researcher and author

[2] Provided E. coli strains encoded with genes for GUS and fusions

[3] Adviser of the primary author and author for correspondence

expressing the wild-type GUS was also treated by the same chromatographic procedure. Behaviors of the transgenic and spiked versions were equivalent. As to position in the elution profile, GUSD0 was eluted at the first target site, and with the additional charges, the fusions were moved to the second site, which provided better selectivity for protein recovery. If one assumes binding through a single fusion (the protein is a tetramer), there is a nearly linear shift in elution within the salt gradient of 17 mM per added charge up to 10 with a reduced increment from 10 to 15. An additional benefit to the strategy of charge manipulation in transgenic canola is that the fusions and their enzymatic activity proved very stable in the extracts through seven days in cold storage, providing flexibility in process scheduling.

Introduction

Plants are alternative hosts to microbial systems for recombinant protein production. Many proteins have been expressed in plants including tobacco (Pen et al., 1992; Verwoed et al., 1995), corn (Hood et al., 1997), potato (Wilmitzer et al., 1992; Sijmons et al., 1990), canola (Krebbers and van de Kerckhove, 1990; Vandekerckhove et al., 1989; Pharmenter et al., 1995), and alfalfa (Austin et al., 1994). Expressed proteins include both pharmaceutical proteins (erythropoietin, Matsumoto et al., 1993) and industrial proteins (phytase, Verwoed et al., 1995), with sizes ranging from 600 (malarial epitopes, Turpen et al., 1995) to 80,000 Da per subunit (phytase, Verwoed et al., 1995) (for more complete list, see Kusnadi et al., 1997). However, there are still two challenges standing in the way of using transgenic plants to produce large quantities of recombinant proteins, namely, to increase the protein

expression level and to recover the expressed protein effectively in downstream processing. The latter problem is addressed here.

Protein recovery from transgenic plants is generally considered difficult because of the large quantity of biomass to be handled (Kusnadi et al., 1997; Whitelam et al., 1993; Whitelam, 1995). Concentration and purification in the initial steps would reduce the recovery costs significantly (Goddijn and Pen, 1995). A proper choice of host can simplify the recovery if the native proteins have very different properties from the recombinant protein. Additionally, purification fusions can add control of the recombinant protein properties to ensure a favorable match with a host (Zhang and Glatz, 1999; Flaschel and Friehs, 1993; Whitelam et al., 1993). Previous studies have shown that properly designed recombinant proteins can facilitate their purification from plants in down steam processing. van Rooijen and Moloney (1995) used oleosins (oil body proteins) as fusions to target GUS to canola oil body to separate GUS from the majority of the seed proteins by a single flotation centrifugation step (13-fold enrichment), and Zhang and Glatz (1999) demonstrated that with a properly controlled charge, T4 lysozyme could be purified by a single step cation exchange chromatography from canola extract with high purity (~90%).

Canola meets the following requirements to be a host for recombinant protein production: seeds or tubers as the protein storage organs to allow long periods of storage (Kusnadi et al., 1997; Austin et al., 1994; Whitelam, 1995; Fiedler and Conrad, 1995); existing large scale production for non-proteinaceous products - starch or oil (Whitelam et al., 1993); low level of toxic components, such as alkaloids, especially for immunotherapeutical and pharmaceutical proteins (Moffat, 1995; Goddijn and Pen, 1995); and known transformation systems.

This paper continues our work of examining canola from the perspective of ease of protein recovery. As reviewed by Zhang and Glatz (1999), canola has a simple storage protein composition. The two major storage proteins, 2S and 12S, have isoelectric point of 11 and 7, respectively (Bhatty et al., 1968; Schwenke, 1994; Lönnerdal and Janson, 1972; Schwenke et al., 1981), and the 12S protein may undergo dissociation into 7S components (molecular weight ~ 150,000 Da) at low ionic strength ($\mu<0.5$) and further into 2-3S protein at low or extremely high pH with the presence of urea (Plietz et al., 1983; Bhatty et al., 1968; Schwenke and Linow, 1982; Schwenke et al., 1983). The nature of the native canola proteins offers an opportunity to produce acidic proteins with pI values less than 7. Presumably, those neutral and basic canola storage proteins would not bind or not bind strongly to an anion exchanger at pH 7.0, allowing a rather simple downstream process for recovering a negatively-charged recombinant protein.

Protein instability during separation is a potential problem. As heterologous proteins, recombinant proteins are susceptible to proteolytic degradation (Flaschel and Friehs, 1993). Polyarginine fusions have shown degradation within the polyarginine portion in canola extracts (Zhang and Glatz, 1999). Highly oxidizing components in some plant systems can destroy enzymatic activity (Jefferson and Wilson, 1991), while other components (such as phenolic acids and condensed tannins) may form complexes with proteins and inhibit enzyme activity (Naczk, et al., 1998; Karamac and Amarowicz, 1997; Loomis, 1974; Naczk, et al., 1994; Goldstein and Swain, 1965) or result in protein precipitation (Hagerman and Butler, 1978).

β-Glucuronidase (GUS) is chosen as the model target protein in this work. It is well characterized, but more importantly, it is absent from most higher plants (such as wheat,

tobacco, potato, and canola) to allow easy identification by enzymatic assays (Jefferson et al., 1986; Jefferson et al., 1987; Jefferson and Wilson, 1991). GUS is also very stable, and its enzymatic reaction can be carried out under a broad range of ionic conditions and pH's (Jefferson and Wilson, 1991). In addition, transgenic canola seed carrying the wild-type GUS was available.

In this paper, the elution behavior of native canola proteins on several anion exchangers is examined to look for where in the profile proteins are absent. For one exchanger, the GUS charged fusions are spiked into non-transgenic canola protein extract to assess the ability to target elution with fusion design. Experiments with protein extract obtained from transgenic canola with recombinant GUS are used as controls to assess the validity of the spiking approach, which eliminates the lengthy and costly procedure of producing quantities of several types of transgenic seed.

Materials and Methods

GUS gene construction in *E. coli*

The wild-type *E. coli* GUS coding region containing no carboxy terminal aspartate residues (GUSD0) was cloned as an NcoI/HpaI fragment into the NcoI/XhoI (blunt) sites of the *E. coli* expression vector pET15b (Novagen, Madison, WI). This eliminated the six histidine fusion tag found on pET15b. An oligonucleotide coding for the polyaspartate tail (either 5, 10, or 15) was fused by PCR amplification to the 3' end of the GUS open reading frame immediately upstream of the termination codon. The resulting PCR products were ligated into the pCR-2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's recommended protocol and confirmed by sequencing. The correct D5, D10,

and D15 fragments were subcloned back into the GUSD0/pET15b construct resulting in the full-length variants (named as GUSD0, GUSD5, GUSD10, and GUSD15 according to the number of aspartates on the fusion tail, respectively). These constructs were confirmed by sequencing and transformed into BL21(DE3) (Novagen, Madison, WI) cell lines for expression.

GUS and its production and purification

Estimated from its amino acid composition, the *E. coli* GUS has a pI of about 5.5. It is a tetramer, and each of the monomers has a molecular weight of ~68000 (Jefferson et al., 1986). The estimated net charge for each fusion at pH 7.0 is summarized in Table 3.1.

The *E. coli* strains coded for GUSD0 and its fusions were cultured in LBH media. GUS expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG). Harvested cells were resuspended in buffer containing 10mM Tris-HCl, 1mM EDTA, and 10mM 2-mercaptoethanol (pH 7.0) (Blanco and Nemoz, 1987), and disrupted by sonication after the addition of phenylmethylsulfonyl fluoride (PMSF, 2mM, Jefferson and Wilson, 1991) to inhibit *E. coli* proteases.

Table 3.1. The estimated charge of GUS and its fusions.

Mutants	No. of aspartates in tail	Estimated charge at pH 7.0
GUSD0	0	-75
GUSD5	5	-95
GUSD10	10	-115
GUSD15	15	-135

GUS was purified from the *E. coli* cell extract by a combination of affinity and anion exchange chromatography. Among several reported affinity absorbents that had been used in GUS purification (saccharolactone, *o*-aminophenyl-β-D-glucuronide (Harris et al., 1973), and thiophenyl-β-D-glucuronide (Blanco and Nemoz, 1987)), saccharolactone was selected because of its commercial availability. For GUSD5, GUSD10, and GUSD15, two steps of affinity chromatography (saccharolactone immobilized on cross-linked 4% beaded agarose, Sigma, St. Louis, MO) and one more step of strong anion exchange chromatography (Q-sepharose fast flow, Sigma, St. Louis, MO) were used. In the case of GUSD0, weak anion exchange chromatography (DEAE-sepharose fast flow, Sigma, St. Louis, MO) was used to replace the strong anion exchanger because the latter could not separate GUSD0 from several contaminants. The final recoveries for all the proteins were close to 50%, and SDS-PAGE showed no significant contaminant proteins.

For affinity chromatography, 10 mM Tris-HCl, 10 mM 2-mercaptoethanol, 1 mM EDTA, pH 7.0 was used as the equilibration buffer, and the high salt elution buffer was equilibration buffer with 0.5 M NaCl. A linear gradient elution was used for the first affinity chromatography, and mixed step and gradient elution were used for the second affinity chromatography. For anion exchange chromatography, 50 mM NaPi pH 7.0 was used as the equilibration buffer, and the high salt elution buffer was 50 mM NaPi, 1 M NaCl pH 7.0. GUS fusions were purified by simple linear gradient elution, and mixed step and gradient elution were used in order to obtain purest possible GUSD0.

Protein assays

The GUS assay is based on its ability to hydrolyze *p*-nitrophenyl β-D-glucosiduronic acid (PNPG) to release chromophore *p*-nitrophenol (Kato et al., 1960; Jefferson and Wilson,

1991). GUS activity is expressed as unit/mL, and one unit of GUS liberates 1 nmol *p*-nitrophenol/min from PNPG at 37 °C and pH 7.0. The assay protocol is developed based on the method of Jefferson and Wilson (1991), which used 1 mM PNPG in extraction GUS extraction buffer as assay buffer. The reaction buffer is 50 mM NaPi pH 7.0 containing 10 mM 2-mercaptoethanol and 1 mM EDTA, and the substrate is 80 mM PNPG in 50 mM NaPi pH 7.0. To measure a sample's GUS activity, 950 μL of the reaction buffer, 25 μL of the substrate, and 25 μL of the sample are mixed together in a cuvette, and the cuvette is kept at 37 °C throughout the assay and monitored at 405 nm to obtain the reaction rate for activity calculation.

The concentration of protein was obtained by Bio-Rad protein assay (Bio-Rad, Hercules, CA), with bovine serum albumin (BSA) as the standard.

Canola protein extract

Both non-transgenic and transgenic canola seed were provided by Pioneer Hi-Bred International (Johnston, IA). Non-transgenic canola seed was ground, dehulled, and flaked before oil extraction, and the defatted meal was obtained by two stages, 30 min each, of *n*-hexane extraction (1 g seed/20 mL of hexane) at room temperature (Zhang and Glatz, 1999). The final oil content was < 2% (w/w). Transgenic canola meal was flaked and dehulled, and oil extraction was carried out by *n*-hexane at 60 °C for six stage cross current operation (6 min each stage) with a ratio of 1 g of seed /3 mL of solvent. The final oil content was ~4% (w/w).

Canola extract was obtained by mixing 50 mM NaPi (pH 7.0) with defatted canola meal at 1:10 (w/v) for 30 min (Zhang and Glatz, 1999). The extract protein concentrations

were 23 ± 2 mg/mL and 15 ± 2 mg/mL for non-transgenic and transgenic canola, respectively.

Anion exchange chromatography

Chromatographic experiments were carried out by using a fast performance liquid chromatography (FPLC) system controlled by BioLogic software (Bio-Rad, Hercules, CA). The equilibrating buffer was 50 mM NaPi pH 7.0 (Buffer A) and the high salt elution buffer was 50 mM NaPi, 1 M NaCl pH 7.0 (Buffer B) for all resins. Except where noted, Q-sepharose fast flow (Sigma, St. Louis, MO) was used as the ion exchanger for canola experiments, and 5 ml of sample was loaded onto the column by an injection loop. For spiked samples, 3 mL of canola protein extract was mixed with 2 mL of GUS stock solution prior to injection (0.4 - 1 wt% of total protein as GUS). Protein elution was accomplished by applying a linear gradient from buffer A to buffer B with a slope of 2 mM NaCl/min. 1 mL/min flow rate was adapted for all chromatographic experiments, and fractions were collected for later GUS activity assay. The schematic experimental procedures were illustrated in Zhang and Glatz (1998).

The columns were packed to a bed height of 6.5 ± 0.2 cm in 1 × 10 Econo columns (Bio-Rad, Hercules, CA). Regeneration was by application of 40 mL of 0.5 M NaOH followed by 40 mL of buffer B, and washing to baseline by buffer A before the next chromatographic run.

GUS stability in canola extract

Samples were prepared by mixing canola protein extract with GUS stock solution at a 3:2 ratio (v/v). The samples were kept at 4 °C for 7 days, and the GUS activity was measured daily. After seven days, samples of the stored material were loaded to the Q-

sepharose column and eluted as above. GUS activity assays were performed on collected fractions to identify the elution of GUS and its fusions.

Chemicals and materials

Sodium phosphate and EDTA were purchased from Fisher (Itasca, IL). The dye reagent for protein assay and the gels (4-15% Tris-HCl gradient) for polyacrylamide electrophoresis were purchased from Bio-Rad (Hercules, CA). All anion exchange resins, affinity resin (Saccharolactone immobilized on cross-linked 4% beaded agarose), BSA standard, and other chemicals were purchased from Sigma (St. Louis, MO).

Results and Discussion

Canola as a host for producing negatively charged recombinant proteins

Figure 3.1 shows that most canola proteins do not bind to the three anion exchange resins as we expected from the basic pI values of the bulk native canola proteins. Those proteins that bound could be eluted by a simple salt gradient. When weak anion exchangers are used (DEAE-sepharose fast flow or DEAE-sepharose cross-linked, 6%), the bound protein is separated into two peaks, and the shoulder on the first peak indicates an unresolved peak. The strong anion exchanger, Q-sepharose fast flow, shows a higher resolving power, and the bound protein is separated into three peaks. The Q-sepharose fast flow was used for all subsequent chromatographic experiments.

The native canola protein profile on Q-sepharose fast flow is replotted in Figure 3.2 to show the extent of "contamination" that would co-elute with any recombinant protein being purified. The mass balance indicates that unbound protein accounts for ca. 90% of total protein loaded to the column, while the first peak eluting in the gradient elution

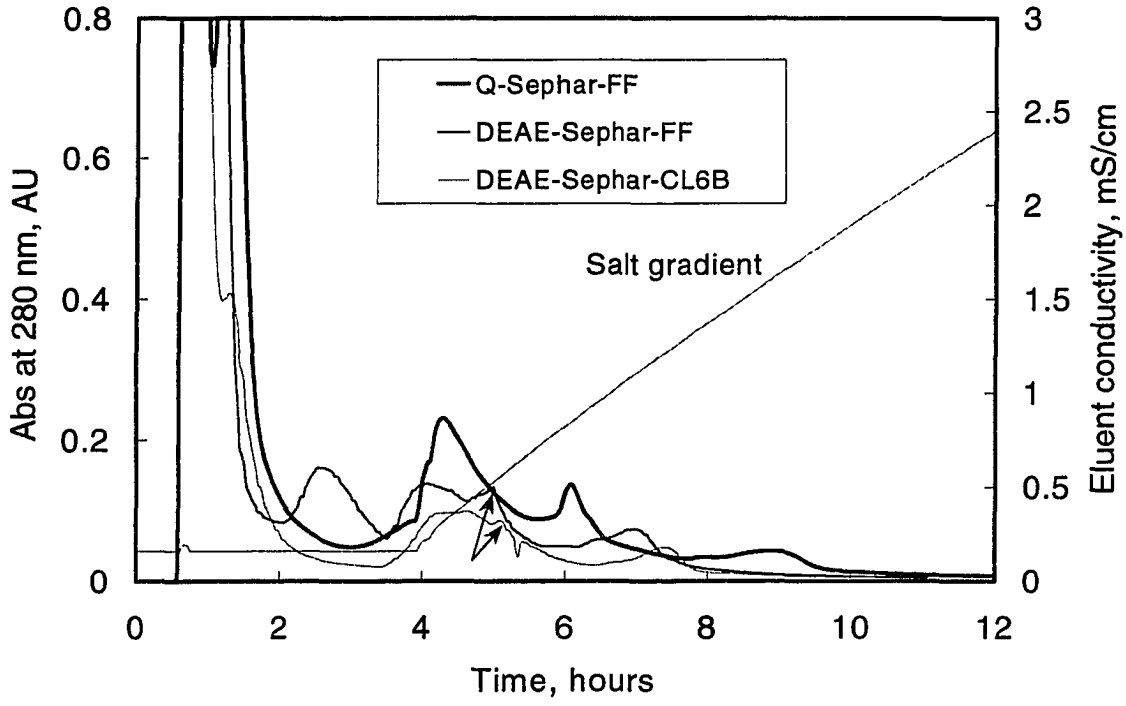


Figure 3.1. Native canola protein elution profiles on different anion exchangers. 6 mL of canola extract was loaded to the column. Gradient elution starts at 3.6 hr (217 min). The two arrows point to the shoulders on the first canola protein peak. Q: functional group is $-CH_2N^+(CH_3)_3$. DEAE: diethylaminoethyl.

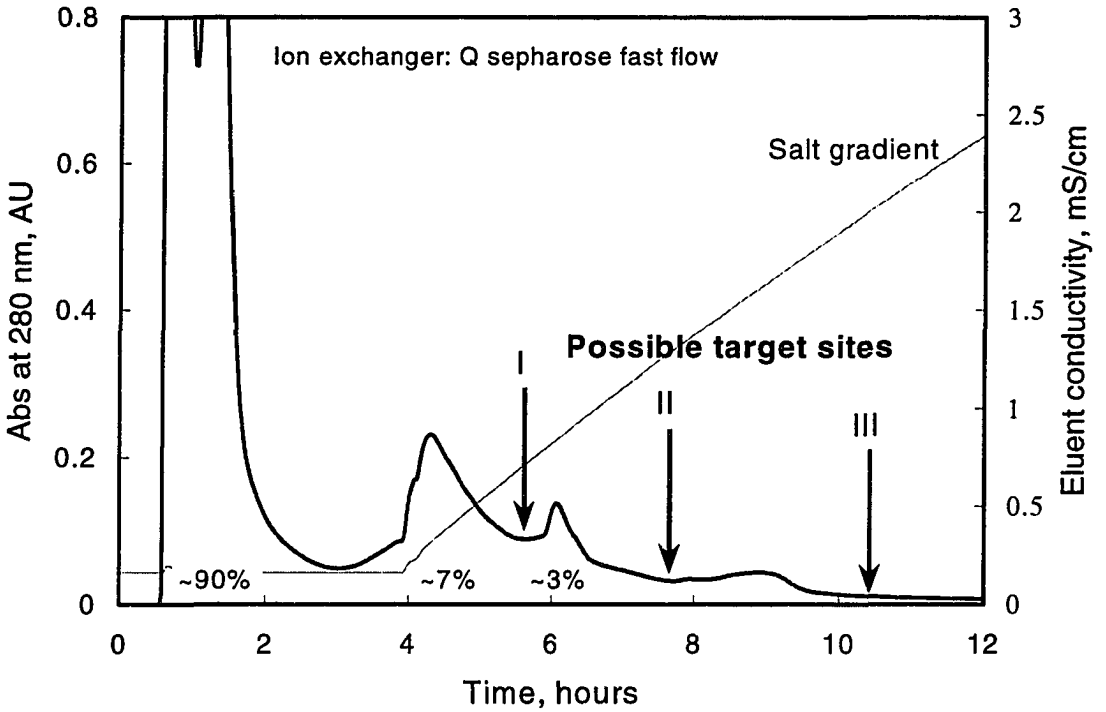


Figure 3.2. Native canola protein elution profile on Q-sepharose fast flow. Arrows point to possible sites to target a genetically engineered protein. Values at the bottom of the corresponding peak indicate the mass percentage of the peak relative to the total loading.

accounts for ca. 7%. The three arrows point to the relatively low background sites where a genetically engineered protein could be targeted to facilitate its purification. Site I is at relatively low salt concentration during gradient elution and has a higher native canola protein background than site II and III. Site III is at very high salt concentration with very low background. It would be the best site to target a protein if possible. Meanwhile, site II provides a compromise between Site I and III. Even if these target sites were not hit, if a recombinant protein were produced in canola and could be eluted after the first bound protein peak, at least 97% of native canola proteins would be eliminated by just one step of anion exchange chromatography. If a protein could be targeted at one of the indicated sites, better purification performance would be expected.

Spiked canola experiment as a model for transgenic canola

The elution profiles (activity peaks) for separate injections of extracts of transgenic canola (expressing GUSD0) and extracts of non-transgenic canola spiked with purified GUSD0 are overlaid in Figure 3.3. The behavior is nearly identical. The slightly earlier elution for the transgenic could be the result of the lower GUS loading for the transgenic (ca. 1000 units, compared to ca. 60000 unites for spiked GUSD0). This gives some basis for the validity of using the spiking experiments to screen potential fusion strategies and hence, reduce the number of candidates for which plant transformants need to be developed.

The purification of GUSD0 and its fusions

The locations where the GUS fusions elute amid the native canola proteins is easily viewed by superimposing the activity assay results from the separate chromatographic runs for the canola samples spiked with each of the fusions in turn (Figure 3.4). The first canola protein peak eluted in Figure 3.4 differs from that of Figure 3.2 in being resolved into two

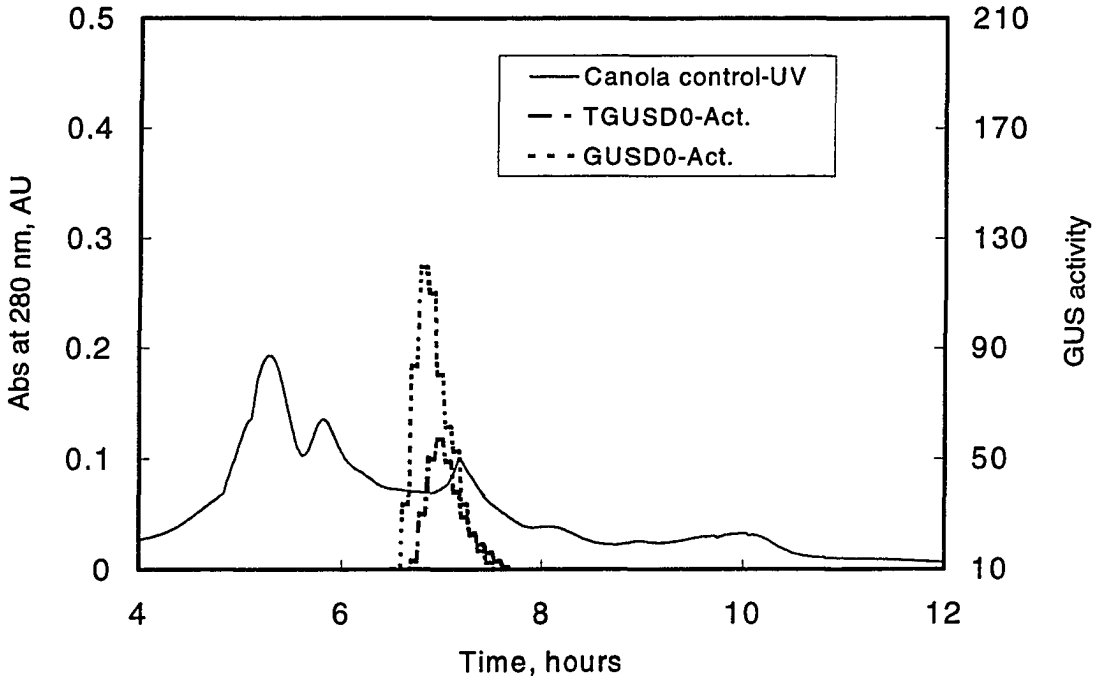


Figure 3.3. Comparison of GUSD0 elution in spiked canola sample and transgenic canola sample (TGUSD0). The activity of spiked GUS has been divided by 15 in order to fit in the same coordinate with that of TGUSD0. Activity unit = u/mL.

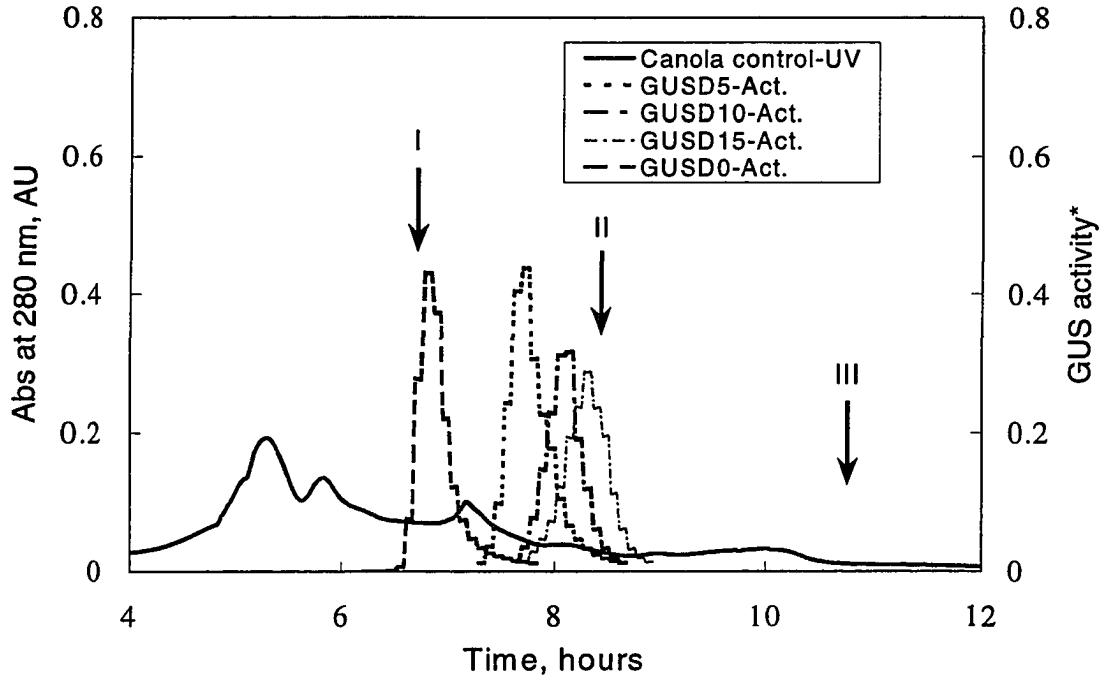
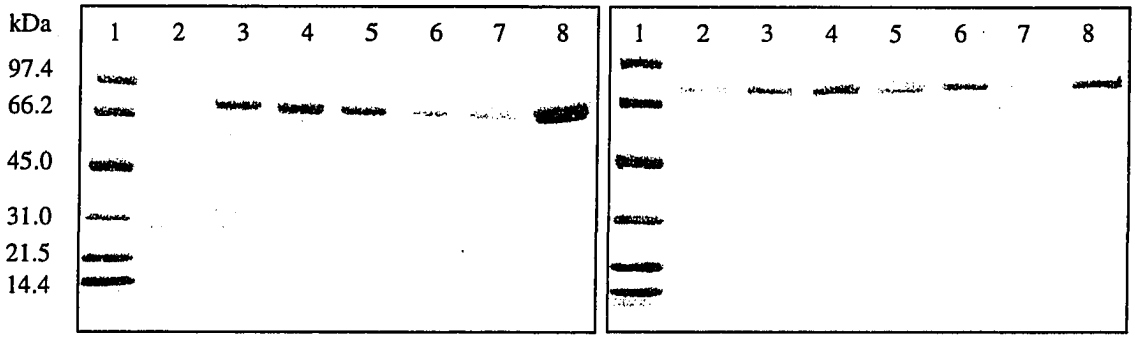


Figure 3.4. Superimpose of GUS activity of different fusions on native canola protein profile. Only part of the profile (4-12 hours) is shown here. 5 mL of spiked samples (for canola control, 3 mL of canola extract + 2 mL of extraction buffer) was loaded. Gradient elution starts at 4.7 hr (280 min). * The actual GUS activity has been multiplied by a factor so all the fusion activity could be fitted in the same window. For GUSD0, actual activity (U/mL) = (value in figure)*5000, for the other three fusions, actual activity (U/mL) = (value in figure)*2500. The arrows point to the possible protein target sites.

peaks. This is because the decreased total protein loading (~50% of the loading in Figure 3.2) improves resolution. GUSD0 is eluted close to site I, and its tail overlaps with the front of the native canola protein peak eluted at higher salt concentration. The addition of poly-aspartate tails moves GUSD5, GUSD10, and GUSD15 to successively longer retention with GUSD15 eluting near site II. The shift on going from 10 to 15 aspartates has a lesser impact on elution than seen in going from 0 to 5 to 10. The reasons for this are being explored via a closer study of the mechanism of binding (Zhang et al., in preparation), but a similar observation has been made for fusions to β-galactosidase (Zhao, et al., 1990; Heng and Glatz, 1995). The trend, however, does imply that site III is out of reach for GUS even with use of longer polyaspartate tails.

The gain in selectivity realized by the addition of tails to reach site II is seen both in the SDS-PAGE gels (Figure 3.5) and the calculated specific activities and enrichment factors (Table 3.2). In Figure 3.5, for GUSD0 (a), fractions have many native canola protein contaminants with molecular weights below 30000 and several other higher molecular weight contaminants (some of them are too faint to be seen). In contrast, for GUSD10 (b), eluting near site II, the contamination level and number of contaminants are significantly decreased. The same conclusion can be reached from the results for final specific activity and enrichment factor shown in Table 3.2. From GUSD0 to GUSD15, the enrichment factor (ratio of final to initial specific activities) has been increased almost four times because the addition of the fusion tail shifts elution to a position with a lower contamination level from native canola proteins. The low yield for GUSD15 is because the activity peak was broader than the six fractions included in the calculations; a much higher yield (~70%) without sacrificing too much on purity would have been reached if more fractions had been included



a) GUSD0-SC b) GUSD10-SC

Figure 3.5. SDS-PAGE for fractions, which show the highest GUS activity, from GUSD0 and GUSD10 spiked canola experiments (GUSD0-SC and GUSD10-SC). Lane 1 shows the low molecular weight markers; Lane 2-7 are the six fractions with highest GUS activity for each experiment; Lane 8 shows corresponding GUS stock solution for each protein. All fractions were concentrated 17 times prior to loading to the gels.

Table 3.2. Summary of GUS purification and recovery based on protein and activity assays.

GUS spiked sample	Initial specific activity, U/mg	Final specific activity, U/mg	Yield %	Enrichment ratio
GUSD0-SC	690	38,000	76.6	55
GUSD5-SC	450	38,000	70.3	84
GUSD10-SC	330	50,000	70.3	152
GUSD15-SC	420	80,000	53.4	190

Total protein loaded to the column in each experiment is 70.5 ± 0.3 mg.

Final specific activity, yield, and enrichment ratio are calculated on the six fractions used in SDS-PAGE experiments.

in the calculation. Consequently, with the addition of negatively charged tails, GUS can be moved from a low salt concentration elution site (site I) to a higher salt concentration elution site (site II) to allow a better purification performance in a single step chromatography.

The effect of additional charges introduced by a tail on the elution behavior of GUS can be seen in Figure 3.6. When the tail length is less than 10, the relationship between the absolute net charge of GUS and the eluent conductivity is approximately linear and unaffected by the presence of canola extract. This proportionality corresponds to a 4.3 mM increase in salt (NaCl) concentration per added charge. Considering the size of GUS and its tetrameric structure, steric effects probably hinder the tails of all four units of the tetramer from binding simultaneously. If, for the sake of rough comparison, we assume that one tail provides for the binding strength, then the increment per added charge is 17.2 mM. Comparison of this value with that calculated on the same basis for charged proteins from

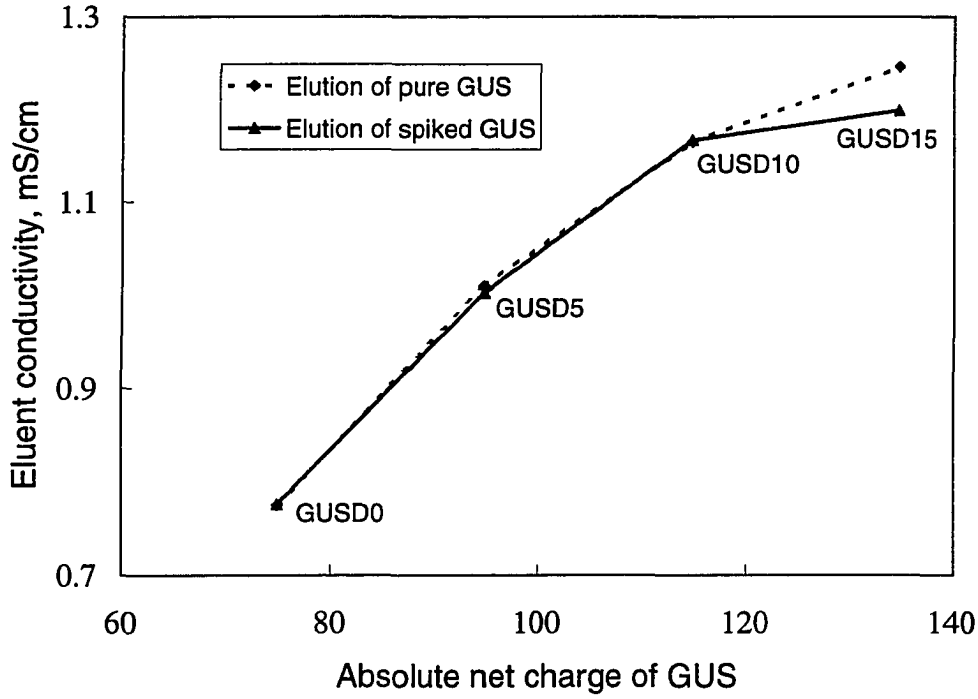
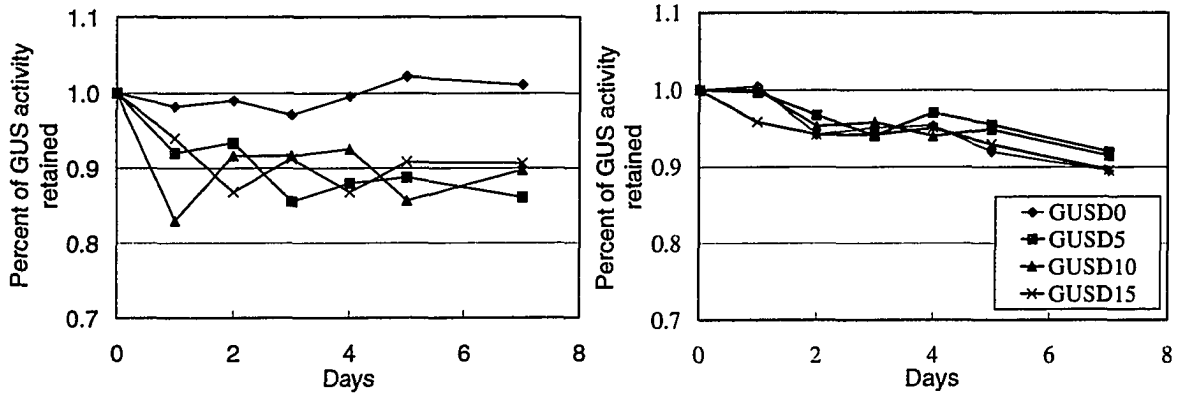


Figure 3.6. Relationship between eluent conductivity and the net charge of GUS. The dotted line indicates the retention of pure GUS samples (GUS stock solution mixed with canola extraction buffer at the same ratio as spiking experiment) under the same experimental conditions.

previous studies (Table 3.3) shows the similar impact of charge additions to the protein retention on different adsorbents.

Enzymatic and fusion stability in canola extracts

Polyarginine tails (positively charged at pH 7.0), fused to T4 lysozyme, were degraded in canola extract as judged by the emergence of multiple peaks during chromatography, which corresponded to a mixture of tail lengths (Zhang and Glatz, 1999). In contrast, all three GUS fusions show a single activity peak in Figure 3.4, indicating that degradation of the fusion tail is not occurring during these experiments. Furthermore, elution of spiked canola samples after 7 days of storage still showed no tail degradation as all the proteins were eluted at the same point (results not shown) as had been seen with fresh



a) GUS controls b) GUS in spiked canola extract

Figure 3.7. GUS stability in canola extract. a) GUS controls. Pure GUS in extraction buffer (50 mM NaPi, 2:3). b) GUS in canola extract. Mixing ratio = GUS stock solution 2 : canola extract 3. All samples were stored in 4 °C refrigerator during this experiment. The GUS activity at day 0 was designated as 1 (100%).

Table 3.3. Comparison of the increment in salt concentration per added charge in protein elution.

Protein	β-glucuronidase	β-galactosidase		T4 lysozyme	human urogastone
Structure	tetramer	tetramer		monomer	polypeptide
MW (monomer)	68 000	116 400		16 800	6 000
Separation method	AEC	AEC	HFIEM	CEC	CEC
Elution method	gradient	gradient /step	step	gradient	gradient
ΔmM NaCl/charge	17.2*	16.8/18.2*	13.6*	27.8	70
Reference	this work	Heng & Glatz, 1995, 1993	Zhao et al., 1990	Zhang &Glatz, 1999	Sassenfeld & Brewer, 1984; Gregiry & Preston, 1977

AEC = anion exchange chromatography.

CEC = cation exchange chromatography.

HFIEM = hollow fiber ion exchange membrane adsorption.

* Values obtained based on the additional charges on a monomer.

samples in Figure 3.4. Polyanionic fusions have also been stable in microbial systems (Ford et al., 1991; Zhao et al., 1990; Parker et al., 1990).

The effect on GUS activity of the storage time in spiked canola extract is compared to that of storage in extraction buffer (50 mM NaPi pH 7.0) in Figure 3.7. After storage at 4 °C for 7 days, GUSD0 and all its fusions still retain 90% of their activity. Comparison to the GUS controls (Figure 3.7, (a)) leads to the conclusion that the slow decline of GUS activity is not because of any detrimental factors (precipitation, protease activity, or inhibition) from canola extract.

Conclusions

This work has demonstrated that canola is suitable as a host for producing acidic recombinant proteins. The simple native canola protein elution profile on anion exchange chromatography provides an opportunity to recover a negatively-charged protein effectively in a single chromatographic step. Negatively-charged fusion tails can enhance a protein's enrichment significantly when the protein does not naturally elute at one of the contaminant-free points of the elution profile. For the case of GUS, a tail with ten or more aspartates can move the GUS from the first target site to the second, providing better purification. The magnitude of the shift for proteins in the extract can be reasonably predicted from the behavior of the pure proteins. The agreement between the spiking and transgenic canola experiments provides preliminary validation of the spiking technique as a cost-effective approach for studying a protein's behavior in the canola system.

Fusion proteins with negatively-charged tails are stable during cold storage in canola extract, both with regards to activity and the fusion tail. This stability would allow for moderate processing delays after the protein extraction.

Acknowledgment

This work was supported by the NSF under Grant No. BES9522644. The authors are very grateful for the kind help from ProdiGene, College Station, TX, for providing us *E. coli* strains encoded with GUS and its fusions. We thank Pioneer Hi-Bred International, Inc., Johnston, IA, for providing the canola seeds. We also thank Dr. Ann Kusnadi for kindly sharing the GUS assay protocol with us and Ms.Yun Bai for providing us defatted transgenic canola meal.

Literature Cited

Austin, S.; Bingham, E. T.; Koegel, R. G.; Mathews, D. E.; Shahan, M. N.; Straub, R. J.; Burgess, R. R. An overview of a feasibility study for the production of industrial enzymes in transgenic alfalfa. *Ann. NY Acad. Sci.* **1994**, *721*, 234-244.

Bhatty, R. S.; McKenzie, S. L.; Finlayson, A. J. The proteins of rapeseed (*Brassica napus L.*) soluble in salt solution. *Can. J. Biochem.* **1968**, *46*, 1191-1197.

Blanco, C.; Nemoz, G. One step purification of *Escherichia coli* β-glucuronidase. *Biochimie.* **1987**, *69*, 157-161.

Fielder, U.; Conrad, U. High-level production and long-term storage of engineered antibodies in transgenic tobacco seeds. *Bio/Technol.* **1995**, *13*, 1090-1093.

Flaschel, E.; Friehs, K. Improvement of downstream processing of recombinant proteins by means of genetic engineering methods. *Biotechnol. Adv.* **1993**, *11*, 31-78.

Ford, C.; Suominen, I.; Glatz, C. E. Fusion tails for the recovery and purification of recombinant proteins. *Protein Expr. Purif.* **1991**, *2*, 95-107.

Goddijn, O. J. M.; Pen, J. Plants as bioreactors. *Trends Biotechnol.* **1995**, *13*, 379-387.

Goldstein, J. L.; Swain, T. The inhibition of enzymes by tannins, *Phytochem.* **1965**, *4*, 185-192.

Gregiry, H.; Preston, B. M. The primary structure of human urogastrone. *Int. J. Pept. Protein Res.* **1977**, *9*, 107-118.

Hagerman, A.; Butler, L. Protein precipitation method for the quantitative determination of tannins. *J. Agr. Food Chem.* **1978**, *26*, 809-812.

Harris, R. G.; Rowe, J. J. M.; Stewart, P.S.; Williams, D. C. Affinity chromatography of β-glucuronidase. *FEBS Lett.* **1973**, *29/2*, 189-192.

Heng, M. H.; Glatz, C. E. Charged fusions for selective recovery of β-galactosidase from cell extract using hollow fiber ion-exchange membrane adsorption. *Biotechnol. Bioeng.* **1993**, *42*, 333-338.

Heng, M. H.; Glatz, C. E. Charged fusions for β-galactosidase retention in anion-exchange chromatography. *J. Chromatogr.* **1995**, *689*, 227-234.

Hood, E. E.; Witcher, D. R.; Maddock, S.; Meyer, T.; Baszcynski, C.; Bailey, M.; Flynn, P.; Register, J.; Marshall, L.; Bond, D.; Kullisek, E.; Kusnadi, A.; Evangelista, R.; Nikolov, Z.; Wooge, C.; Mehigh, R. J.; Hernan, R.; Kappel, W. K.; Ritland, D.; Li, C.-P.; Howard, J. A. Commercial production of Avidin from transgenic maize: Characterization of

transformant, production, processing, extraction and purification. *Mol. Breed.* **1997**, 3, 291-306.

Jefferson, R. A.; Burgess, S. M.; Hirsh, D. β-Glucuronidase from *E. coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8447-8451.

Jefferson, R. A.; Kavanagh, T. A.; Bevan, M. W. GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO J.* **1987**, *6*, 13, 3901-3907.

Jefferson, R. A.; Wilson, K. J. The GUS gene fusion system. *Plant Mol. Biol. Manual* **1991**, *B14*, 1-33.

Karamac, M.; Amarowicz, R. Inhibition of pancreatic lipase by phenolic acid - examination in vitro. *Zeitschrift für Naturforschung* **1997**, *51C*, 903-905.

Kato, K.; Yoshida, K.; Tsukamoto, H.; Nobunaga, M.; Masuya, T.; Sawada, T. Synthesis of *p*-nitrophenyl β-D-glucosiduronic acid and its utilization as a substrate for the assay of β-Glucuronidase activity. *Chem. Pharm. Bull.* **1960**, *8*, 239-242.

Krebbers, E. and van de Kerckhove, J. Production of peptides in plant seeds. *Trends Biotechnol.* **1990**, *8*, 1-3.

Kusnadi, A.; Nikolov, Z. L.; Howard, J. Production of recombinant proteins in transgenic plants: Practical considerations. *Biotechnol. Bioeng.* **1997**, *56*, 473-484.

Loomis, W. Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Methods Enzymol.* **1974**, *31*, 528-544.

Lönnerdal, B.; Janson, J.-C. Studies on Brassica seed proteins. *Biochim. Biophys. Acta* **1972**, *278*, 175-183.

Matsumoto, S.; Ishii, A.; Ikura, K.; Ueda, M.; Saskl, R. Expression of human erythopoietin in cultured tobacco cells. *Biosci. Biotechnol. Biochem.* **1993**, *57*, 1249-1252.

Moffat, A. S. Exploring transgenic plants as a new vaccine source. *Science* **1995**, *268*, 658-660.

Naczk, M.; Amarowicz, R.; Sullivan, A.; Shahidi, F. Current research developments on polyphenolics or rapeseed/canola: a review. *Food Chem.* **1998**, *62*, 489-502.

Naczk, M.; Nichols, T.; Pink, D.; Sosulski, F. Condensed tannins in canola hulls. *J. Agr. Food Chem.* **1994**, *42*, 2196-2200.

Oil crop yearbook, 1995, USDA.

Parker, D. E.; Glatz, C. E.; Ford, C. F.; Gendel, S. M.; Suominen, I.; Rougvie, M. A. Recovery of a charged-fusion protein from cell extracts by polyelectrolyte precipitation. *Biotechnol. Bioeng.* **1990**, *36*, 467-475.

Pen, J.; Molendijik, L.; Quax, W. J.; Sijmons, P. C.; van Ooyen, A. J. J.; van den Elzen, P. J. M.; Rietvelk, K.; Hoekema, A. Production of active *Bacillus licheniformis* alpha-amylase in tobacco and its application in starch liquefaction. *Bio/Technol.* **1992**, *10*, 292-296.

Pharmenter, D. L.; Boothe, J. G.; van Rooijen, G. J. H.; Yeung, E. C.; Moloney, M. M. Reduction of biologically active hirudin in plant seeds using oleosin partitioning. *Plant Mol. Biol.* **1995**, *29*, 1167-1180.

Plietz, P.; Damaschum, G.; Muller, J. J.; Schwenke, K. D. The structure of 11-S globulin from sunflower and rapeseed: A small-angle X-ray scattering study. *Eur. J. Biochem.* **1983**, *130*, 315-320.

Sassenfeld, H. M.; Brewer, S. J. A polypeptide fusion designed for the purification of recombinant proteins. *Bio/Technol.* **1984**, *2*, 76-81.

Schwenke, K. D. In *New and developing sources of food proteins*; Hudson, B. J. F., Ed.; Chapmam & Hall: New York, **1994**; pp 281-306.

Schwenke, K. D.; Linow, K.-J. A reversible dissociation of the 12S globulin from rapeseed (Brassica napus L.) depending on ionic strength. *Nahrung*, **1982**, *26*, K5-K6.

Schwenke, K. D.; Raab, B.; Plietz, P.; Damaschum, G. The structure of the 12S globulin from rapeseed (Brassica napus L.). *Nahrung*, **1983**, *27*, 165-175.

Schwenke, K. D.; Raab, B.; Linow, K.-J.; Pahtz, W.; Uhlig, J. Isolation of the 12S globulin from rapeseed (Brassica napus L.) and characterization as a "neutral" protein. *Nahrung*, **1981**, *25*, 271-280.

Sijmons, P. C.; Dekker, B. M.; Schrammeijier, B.; Verwoerd, T. C.; van den Elzen, P. J. M.; Hoekeman, A. Production of correctly processed human serum albumin in transgenic plants. *Bio/Technol.* **1990**, *8*, 217-221.

Turpen, T. H.; Reinl, S. J.; Charoenvit, Y.; Hoffman, S. L.; Fallarme, V.; Grill, L. K. Malarial epitopes expressed on the surface of recombinant tobacco mosaic virus. *Bio/Technol.* **1995**, *13*, 53-57.

Vandekerckhove, J.; Damme, J. V.; Lijsebettens, M. V.; Botterman, J.; Block, M. D.; Vandewiele, M.; Clercq, A. D.; Leemans, J.; Montagu, M. V.; Krebbers, E. Enkephalins produced in transgenic plants using modified 2S seed storage proteins. *Bio/Technol.* **1989**, *7*, 929-932.

van Rooijen, G. J. H.; Moloney, M. M. Plant seed oil-bodies as carriers for foreign proteins. *Bio/Technol.* **1995**, *13*, 72-77.

Verwoerd, T. C.; van Paridon, P. A.; van Ooyen, A. J. J., van Lent, J. W. M.; Hoekema, A.; Pen, J. Stable accumulation of *Aspergillus niger* phytase in transgenic tobacco leaves. *Plant Physiol.* **1995**, *109*, 1199-1205.

Wilmitzer, L.; Sonnewald, U.; Roeber, M.; Carlsen, S.K. Transgenic plants expressing genes for industrial enzymes. International Patent, **1992**, WO 92/01042.

Whitelam, G. C. The production of recombinant proteins in plants. *J. Sci. Food Agr.* **1995**, *68*, 1-9.

Whitelam, G. C.; Cockburn, B.; Gandecha, A. R.; Owen, M. R. L. Heterologous protein production in transgenic plants. *Biotechnol. Genet. Eng. Rev.* **1993**, *11*, 1-29.

Zhang, C.; Glatz, C. E. Purification of β-glucuronidase and its fusions from spiked canola protein extract by anion exchange chromatography. Reilly, P. J. Ed. The proceedings of the 28th annual biochemical engineering symposium, Iowa State University, Ames, Iowa, **1998**, pp17-22.

Zhang, C.; Glatz, C. E. Processing engineering strategy for recombinant protein recovery from canola by cation exchange chromatography. *Biotechnol. Prog.* **1999**, 15, 12-18.

Zhang, C.; Venkatesh, N.; Cramer, S.; Glatz, C. E. Prediction of the elution of genetically engineered proteins in gradient elution chromatography by their isocratic elution characteristics, in preparation.

Zhao, J.; Ford, C. F.; Glatz, C. E.; Rougvie, M. A.; Gendel, S. M. Polyelectolyte precipitation of β-galactosidase fusions containing poly-aspartic acid tails. *J. Biotechnol.* **1990**, *14*, 273-284.

CHAPTER 4. PREDICTION OF GENETICALLY ENGINEERED PROTEIN RETENTION IN LINEAR GRADIENT ELUTION CHROMATOGRAPHY

A paper to be submitted to the Journal of Chromatography

Chenming Zhang[1], Venkatesh Natarajan[2], Steven M. Cramer[2], and Charles E. Glatz[3]

Abstract

Genetically engineered proteins are characterized by the stoichiometric displacement model parameters, Z and I. The model proteins are wild type β-glucuronidase and its fusions. Three fusions, named as GUSD5, GUSD10, and GUSD15, were developed with 5, 10 and 15 aspartic acids on the C-terminal of wild-type GUS. The protein specific parameters, Z and I, were used in equilibrium-based models for protein elution providing a numerical simulation of the elution profile that incorporates dispersion or an algebraic expression for retention time without inclusion of dispersion effects. Both methods were sufficient for protein retention prediction in linear gradient elution, while the simulation provided more detailed information on protein peak shape. All peaks sharpened and shifted to higher eluent salt concentration as the gradient became steeper. The equilibrium binding parameters provided additional information on the function of the fusion tail, and GUSD15 showed different trends in Z and I from the shorter fusions.

---

[1] Graduate student, primary researcher and author

[2] Conducted computer simulation using lumped dispersion model

[3] Adviser of the primary author and author for correspondence

Introduction

Ion exchange chromatography (IEC) is widely used in separation and fractionation of proteins (Bonnerjea, et al., 1986; Morris and Morris, 1976). The most widely used elution methods are isocratic and gradient elution. In isocratic elution, the elution buffer composition is constant during the chromatographic process. On the other hand, gradient elution (mostly linear gradient) involves continuous salt concentration change in the mobile phase, and it is most commonly used for separating proteins that bind to the stationary phase (Yamamoto et al., 1983a, 1987). The ability to predict the elution of an absorbed protein during gradient elution is essential in designing an effective process for protein recovery.

Many models have been proposed to describe protein retention. For isocratic elution chromatography, three theories - equilibrium theory, rate theory (mass balance model) and plate theory (Yamamoto, et al., 1988) - have been widely used to describe protein retention and column efficiency. However, two other models based on mass action - the stoichiometric displacement model and the steric mass action model - provide a practical means of extracting the protein specific parameters for predicting protein retention in gradient elution chromatography. The stoichiometric displacement model (SDM) was introduced and applied to ion-exchange chromatography by Regnier and coworkers (Kopaciewicz, et al., 1983; Rounds and Regnier, 1984; Drager and Regnier, 1986, 1987) based on the work by Boardman and Partridge (1955), and the steric mass action (SMA) model was proposed later by Cramer and coworkers (Brooks and Cramer, 1992; Gerstner, et al., 1994). In the SDM, a bound protein is displaced by certain number of counter ions during elution, and the protein retention is described by two protein specific parameters, characteristic charge and binding affinity. The SMA model additionally takes into

consideration the additional steric obstruction of sites by large biopolymers through introduction of a third parameter – the protein steric factor. Both models have been successfully used in interpreting protein retention and extracting protein specific parameters. The SDM has been widely used in both analytical and preparative mode of ion-exchange chromatography (Round and Regnier, 1984; Drager and Regnier, 1986, 1987; Regnier and Mazsaroff, 1987; Velayudhan and Horvath, 1988, 1994; Heng and Glatz, 1995), but its application to preparative systems has been difficult (Regnier and Mazsaroff, 1987). However, the SMA model has been successfully applied to preparative of ion-exchange (Gallant, et al., 1996a, 1996b).

For gradient elution, many models have been proposed to predict the dependence of the elution profile (the peak position, the peak width and even the peak shape) on the operating conditions (the gradient slope, buffer systems, the flow rate, and others). Those models can be categorized into three groups based on how a protein retention process is considered. One class treats the column as a number of equivalent theoretical plates with peak spreading introduced by protein distribution and diffusion. The second class treats the packed bed as continuous media with peak spreading introduced via rate expressions for mass transfer and reaction steps. The third class does not address the peak spreading but models protein adsorption equilibrium in order to predict elution time. The fundamental assumptions for all the models include the following:

1) The column is homogeneous.
2) The flow distribution is uniform in the column.
3) All processes are isothermal.
4) Concentration gradients in the radial direction of the column are negligible.

Several selected models are listed in Table 4.1. In general, the more rigorous a model is in describing protein retention, the more accurate the model should be able to predict the protein elution. Rate models incorporate the different mass transfer processes (axial dispersion, intraparticle diffusion, and film diffusion) in order to predict peak retention and peak spreading, but require more detailed information about the column and protein for the numerical computation (Berninger, et al., 1991; Ma and Guiochon, 1991). In plate models, zone spreading effects are represented by the number of plates, but the measurement of parameters, such as protein distribution and diffusion coefficient, which are necessary for calculating the plate number, is rather laborious (Yamamoto, et al., 1983a, 1983b). On the

Table 4.1. Comparison of different models used in predicting protein retention in ion-exchange chromatography.

Model	Mass transport considered	References
Plate model	Axial dispersion; intraparticle diffusion	Yamamoto et al. 1983a, 1983b
Rate model (VERSE-LC)	Axial dispersion; intraparticle diffusion; film mass transfer	Berninger, et al. 1991
Rate model (Lumped Dispersion model)	Axial dispersion; intraparticel diffusion	Ma and Guiochon, 1991; Rhee and Amundson, 1972
Mass action model	None	Stout, et al. 1986
Mass action model	None	Jandera and Churacek, 1974a, 1974b

VERSE-LC: VErsatile Reaction-SEparation model for Liquid Chromatography.

other hand, models based upon mass action laws (Snyder, 1964; Snyder, et al., 1979; Stout, et al., 1986; Jandera and Churacek, 1974a, 1974b) ignore zone spreading effects and are straightforward to use but predict only peak retention.

An additional factor to be accounted for is that the operating conditions (the gradient slope, buffer systems, flow rate) affect the protein elution (Yamamoto, et al., 1983a, 1983b). The protein peak position and peak width depend on the slope of the gradient in ion-exchange chromatography (Novotny, 1971; Kawasaki and Bernardi, 1970; Kato et al., 1982; Luo and Hsu, 1997).

The development of recombinant DNA technology has allowed a protein being genetically engineered to favor its purification produced from both microbial (Sassenfeld and Brewer, 1984) and plant systems (Zhang and Glatz, 1999; Zhang, et al., 1999). SDM has been used to characterize a series of β-galactosidase purification fusions on anion-exchange chromatography (Heng and Glatz, 1995) and subtilisin on cation-exchange chromatography (Chicz and Regnier, 1989), but no effort has been made to simulate the elution of such genetically engineered proteins. It is the goal of this paper to apply the SDM to characterize the equilibrium binding of a series of β-glucuronidase fusions with different length of fusion tails and use that information to predict gradient elution of these proteins.

Theoretical Framework

Stoichiometric displacement model

The stoichiometric displacement model (SDM) (Kopaciewicz, et al., 1983; Drager and Regnier, 1986, 1987) characterizes an ion-exchange process as

$$P_0 + Z \cdot C_b \Leftrightarrow P_b + Z \cdot C_0 \qquad (1)$$

where P_0 and P_b refer to protein concentration at the free state in the mobile phase and at the bound state on the stationary phase, C_b and C_0 refer to the concentration of the bound and free salt ion. Z is the number of bound ions displaced during the protein adsorption process. The equilibrium constant (K_{eq}) for this displacement reaction is expressed as

$$K_{eq} = \frac{P_b \cdot C_0^{\,Z}}{P_0 \cdot C_b^{\,Z}} \tag{2}$$

The protein retention (as the capacity factor k') can be related to the displacing ion concentration in mobile phase (C_0) by

$$k' = \frac{I}{C_0^{\,Z}} \tag{3}$$

where I is a solute specific constant. It indicates the over all (specific and non-specific interactions) affinity of the solute to the sorbent surface (Chicz and Regnier, 1989), and when a column is operated in the analytical mode (the fraction of the surface loaded with protein is negligible), it can be related to the equilibrium constant by (Kopaciewicz, et al., 1983; Jandera and Churacek, 1974a)

$$I = K_{eq}\varphi\Lambda^{Z} \tag{4}$$

where φ is the phase ratio (the ratio of the total volumes of stationary and mobile phase in a column), and Λ is the ion capacity of the resin.

The capacity factor (k') is related to the protein retention by (Drager and Regnier, 1986)

$$k' = \frac{(t_R - t_0)}{t_0} \tag{5}$$

where t_R and t_0 are the retention times of the solute at retained or non-retained conditions, respectively.

The linear form of Eq. (3)

$$\log k' = \log I + Z \log(C_0)^{-1} \tag{6}$$

can be used to obtain the two protein-specific parameters, Z and I, by plotting k' obtained from Eq. (5) at a series of isocratic elution conditions (C_0).

Protein elution simulation

Mass transport equations

The single parameter, lumped dispersion model (Czok and Guiochon, 1990; Ma and Guiochon, 1991) is used to incorporate axial dispersion into the equilibrium model and thus account for peak shape. In a packed bed chromatography column, assuming local equilibrium, the differential mass balance on solute in a mobile phase in plug flow is

$$-\left(\frac{D_i}{u_0 L}\right)\frac{\partial^2 P_0}{\partial x^2} + \frac{\partial P_0}{\partial x} + \frac{\partial P_0}{\partial \tau} + \frac{(1-\varepsilon)}{\varepsilon}\frac{\partial P_b}{\partial \tau} = 0 \tag{7}$$

where D_i is the effective dispersion coefficient, which lumps the effects of dispersion and molecular diffusion; u_0 is the chromatographic velocity ($u_0 = u_l / \varepsilon$, u_l is the linear velocity, cm/s); L is the column length; ε is the total porosity of the column; $x = X / L$ is the dimensionless axial position; $\tau = t / t_0$ (t_0 is the column dead time) is the dimensionless time. P_b is obtained from the SDM equilibrium formalism, Eq. (2).

The initial conditions (when $\tau = 0$) for solving the above partial differential equation are:

$$C_0(0, x) = C_{0,e} \tag{8a}$$

$$C_b(0,x) = \Lambda \tag{8b}$$

$$P_0(0,x) = 0 \tag{8c}$$

$$P_b(0,x) = 0 \tag{8d}$$

where $C_{0,e}$ is the anion concentration of the equilibrating buffer. During the chromatographic operation, the boundary conditions (when $x = 0$) for the displacing ions are:

$$C_0(0 < \tau < \tau_f, 0) = C_{0,e} \qquad \text{(Loading)} \tag{9a}$$

$$C_0(\tau > \tau_f, 0) = C_{0,e} + G \cdot (\tau - \tau_f) \qquad \text{(Gradient)} \tag{9b}$$

where τ_f is the length of the sample feed pulse and G is the gradient slope in dimensionless time scale (mM). For a protein, the boundary conditions are:

$$P_0(0 < \tau < \tau_f, 0) = P_{0,f} \qquad \text{(Loading)} \tag{10a}$$

$$P_0(\tau > \tau_f, 0) = 0 \qquad \text{(Gradient)} \tag{10b}$$

where $P_{0,f}$ is the protein concentration in a feed sample.

Solution of the mass transfer equation

The observed effective dispersion coefficient (D_i) in Eq. (7) is obtained from (Czok and Guiochon, 1990; Ma and Guiochon, 1991)

$$\frac{H}{L} = \frac{2D_i}{Lu_0} \tag{11}$$

by plotting the height equivalent to a theoretical plate (H) under various chromatographic velocities under conditions where the solute is not retained. In doing so, the velocities are chosen so that the elution peaks are as symmetric as possible.

Equation (7) was solved by employing a finite difference scheme (Gallant, et al., 1996), and simulated chromatograms were generated with a computer program written in FORTRAN.

Equation based on mass action mechanism for predicting protein retention

The equation used in predicting protein retention based on the mass action mechanism was derived by Jandera and Churacek (1974a, 1974b). During gradient elution with a monovalent salt in anion-exchange chromatography, the displacing anion concentration, C_0, at which a protein peak elutes, can be calculated by

$$C_0 = \left\{ \left[G'(\chi \cdot Z + 1) \cdot I \cdot V_m + C_{0,e}^{\frac{\chi \cdot Z+1}{\chi}} \right]^{\frac{1}{\chi \cdot Z+1}} - G' V_d \right\}^{\chi} \qquad (12)$$

where χ is the shape factor for the concentration gradient. The concentration gradient is concave for $\chi > 1$, convex for $\chi < 1$, and linear for $\chi = 1$. V_m is the total volume of the mobile phase in the column; V_d is the dead volume of the connection tube between the outlet of the gradient generating device and the top of the column; G' is the gradient slope in mobile phase volume scale (mM/mL). All the other symbols have described earlier.

For a linear gradient elution operation, Eq. (12) becomes

$$C_0 = \left[G'(Z+1) \cdot I \cdot V_m + C_{0,e}^{Z+1} \right]^{\frac{1}{Z+1}} - G' V_d \qquad (13)$$

In Eq. (13), Z and I are protein specific parameters obtained from the SDM. V_m and V_d are directly measurable, and gradient slope, G', is the control parameter. Hence, knowledge of Z and I is sufficient to calculate protein retention at a given gradient.

On the other hand, the lumped dispersion model (Eq. 2 and 7) needs other parameters, namely the column-specific parameters, binding capacity and porosity, and the protein

dispersion coefficient in addition to Z and I. The additional information gained on the elution profile is important for determination of resolution of multiple components.

Material and Methods

Materials

Sodium monobasic phosphate, sodium dibasic phosphate, and sodium chloride were purchased from Fisher (Itasca, IL). The econo-column, Q-sepharose fast flow resin, BSA standard, and other chemicals were purchased from Sigma (St. Louis, MO).

GUS and its fusions

β-Glucuronidase (GUS) of *E. coli* is very stable and has a pI of about 5.5. The active protein is a tetramer, and each monomers has a molecular weight of ~68000 (Jefferson, et al., 1986). Three fusions were developed based on the wild-type GUS, and they are designated as GUSD0, GUSD5, GUSD10, and GUSD15, according to the number of aspartates on the fusion tail. The estimated charge for each fusion at pH 7.0 is -75, -95, -115, and -135, respectively (Zhang, et al., 1999).

GUSD0 and its fusions were produced by fermentation of engineered *E. coli* strains provided by ProdiGene (College Station, TX), and the gene expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG). The protein purification was accomplished by using combined affinity and anion-exchange chromatography (Zhang, et al., 1999).

The GUS assay is based on protein's ability to hydrolyze *p*-nitrophenyl β-D-glucosiduronic acid (PNPG) to release chromophore *p*-nitrophenol (Kato et al., 1960;

Jefferson and Wilson, 1991), and GUS activity is expressed as unit/ml. One unit of GUS can liberate 1 nmol *p*-nitrophenol/min from PNPG at 37 °C and pH 7.0 (Zhang, et al., 1999).

The protein concentrations are determined by Bio-Rad protein assay (Bio-Rad, Hercules, CA), with bovine serum albumin (BSA) as the standard.

Chromatography

The strong anion exchanger, Q-sepharose fast flow (bead average diameter ~90 μm), was packed into a 15 × 1 cm I.D. empty econo-column to a final height of 6.7 cm (5.26 mL). A flow adaptor was used to eliminate the dead volume on the top of the packed bed and ensure even distribution of sample.

Chromatographic experiments were carried out by using a fast performance liquid chromatography (FPLC) system controlled by BioLogic software (Bio-Rad, Hercules, CA). The equilibrating buffer used was 50 mM NaPi pH 7.0 (Buffer A), and the high salt elution buffer was 50 mM NaPi, 1 M NaCl, pH 7.0 (Buffer B). Fractions of the column effluent were collected and assayed to identify the GUS peak. All experiments were repeated at least twice.

System parameters

The dead volume from the outlet of the gradient generation device to the top of the bed, V_d, was determined by directly measuring the volume of the connection tubing.

The column total porosity was determined from the retention time by injecting a pulse of acetone in Buffer A (0.1 mL, 10% v/v) into the column, which was eluted at 0.1 mL/min by 50 mM NaPi pH 7.0 and monitored at 280 nm (Kim, et al., 1992).

The bed ion capacity, Λ, was determined by frontal analysis according to Gadam et al. (1993). The column was first equilibrated by 50 mM Tris-HCl pH 7.0. A front of

different concentration of sodium nitrate in 50 mM Tris-HCl pH 7.0 was then introduced at 1 mL/min, and the eluent was monitored at 280 nm. The capacity was determined from the nitrate breakthrough front.

Results and Discussion

Column parameters

Column parameters measured by the methods described above are: $V_d = 0.61$ mL; $\varepsilon = 0.903$; $\Lambda = 201$ mM.

Protein parameters, Z and I, in SDM

The SDM parameters for each protein are determined by measurement of capacity factors, k', for a series of isocratic elution experiments carried out at different salt concentrations, C_0, and the results plotted according to Eq. (6) in Figure 4.1. The 50 mM phosphate anion in the equilibrating buffer is estimated to be equivalent to 50 mM of chloride anion. Values of characteristic charge (Z) and the solute specific parameter (I) from such plots, along with the estimated protein net charges at neutral pH are summarized in Table 4.2.

The characteristic charge of GUS increases 4 units with additional 5 aspartates on the tail for GUSD0, GUSD10, and GUSD10. A simplistic interpretation would be that one of the fusion tails on the tetramer is able to bind in addition to the sites distributed over the native GUS. K_{eq} increases as well. However, the increment from 10 to 15 aspartates results in not only a smaller Z, but also a smaller equilibrium constant. The decreased effectiveness of binding behavior of the longer-tailed GUSD15 is consistent with previously reported work

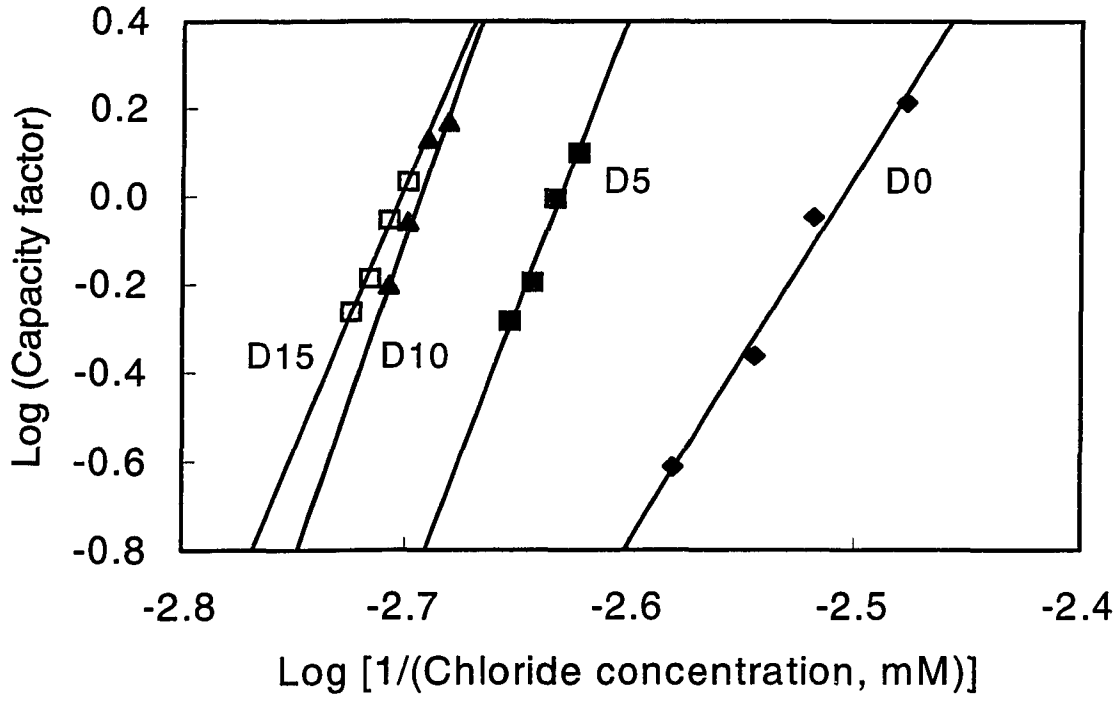


Figure 4.1. Plot of capacity factor versus chloride concentration for GUS and its fusions based on the stoichiometric displacement model. The protein names have been abbreviated as D0, D5, D10, and D15, respectively.

Table 4.2. Estimated and SDM characteristic charges and protein specific parameters.

Protein	Estimated net charge	Characteristic charge Z	Protein parameter I	Equilibrium constant K_{eq}*
GUSD0	-75	8.3	5.2×10^{20}	4.4×10^{2}
GUSD5	-95	12.2	1.5×10^{32}	9.7×10^{4}
GUSD10	-115	16.1	3.3×10^{43}	2.1×10^{7}
GUSD15	-135	12.1	5.1×10^{32}	6.3×10^{5}

* Calculated from Eq. (4). $\varphi = (1-\varepsilon)/\varepsilon = 0.107$.

done on β-galactosidase with various length of aspartate tails (Heng and Glatz, 1995; Niederauer, et al, 1994). The cause may be an interaction of the tail with the core GUS or perhaps the length promotes secondary structure formation. Drager and Regnier (1986) proposed the latter cause in explaining the decreased Z values in oilgodeoxynucleotides.

Protein elution simulation

Effective dispersion coefficients, D_i

The effective dispersion coefficients, D_i, are obtained from Eq. (12) based on a series of isocratic elution experiments when proteins are at non-retained conditions. The flow rates were chosen as low as possible to yield symmetric elution peaks, and the injection volume was 100 μL. The results are plotted in Figure 4.2, and the dispersion coefficients, calculated from the peak widths, and the protein loading concentrations used in the simulation are listed in Table 4.3. As expected, all the fusions have a similar dispersion coefficient, since the

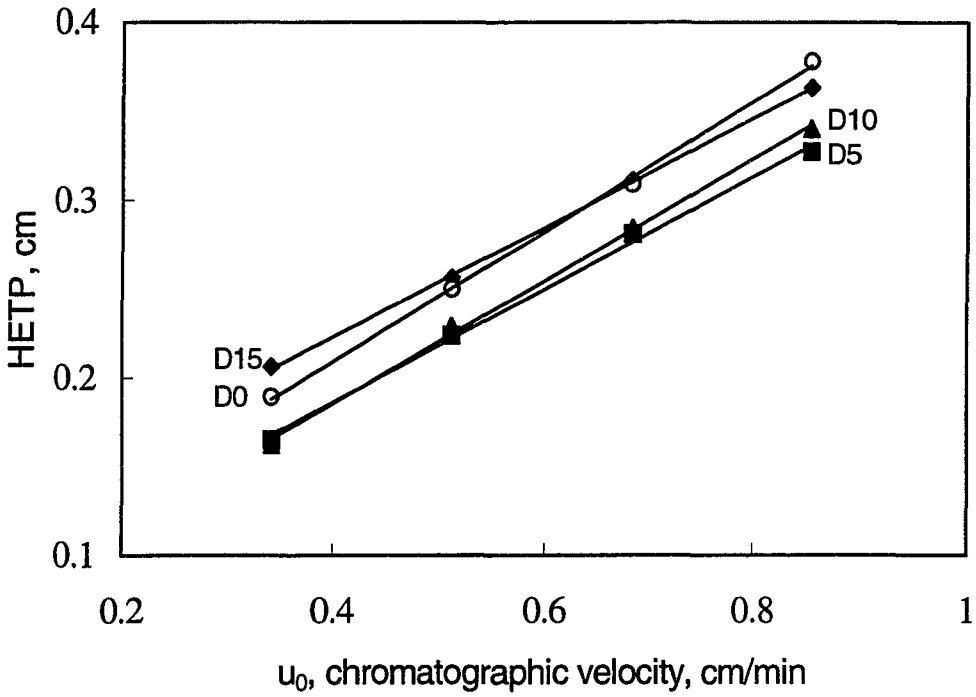


Figure 4.2. Plot of HETP versus the chromatographic velocity, u_0, for GUS and its fusions. Equilibrating buffer and sample carrier buffer was 50 mM NaPi, 2 M NaCl, pH 7.0; Injection volume = 100 µL.

Table 4.3. Protein dispersion coefficient (D_i) and protein loading concentration ($P_{0,f}$) used in elution simulation.

Protein	D_i, cm^2/min	$P_{0,f}$, mM
GUSD0	0.15	8.64×10^{-4}
GUSD5	0.13	9.68×10^{-4}
GUSD10	0.16	6.52×10^{-4}
GUSD15	0.18	7.92×10^{-4}

column is the same, the proteins are of nearly the same size, and solution conditions are the same.

Comparison of simulation to experimental results

Linear gradient experiments with four different gradient slopes were carried out in order to examine the simulation's ability to capture the effect of changing elution conditions.

For each experiment, 250 μL of single component protein sample (concentrations are listed in Table 4.3, sample buffer: 50 mM NaPi, pH 7.0) was injected into the column. After a short period of washing by equilibrating buffer (50 mM NaPi, pH 7.0), a linear gradient was introduced into the column. UV absorbance was not sufficient for identification of the GUS peak because of a UV absorbing nonprotein coeluting with GUSD0 that also overlapped with the fusions for the higher gradients. Hence, all protein elution peaks were identified by activity assays on collected fractions. The total GUS activity loaded to the column for each protein was ~6000 units. The GUS recovery after the gradient varied from 50% to 90%. To avoid the inconsistency in GUS assay and for the purpose of comparison, the recovery is assumed to be 100%, and each of the activity peaks shown is normalized to a

peak area corresponding to 6000 units. The activity peaks of the four GUS proteins under same gradient elution are superimposed in the same chromatogram, and the comparisons between experimental results and the simulations are shown in Figures 4.3 to 4.6.

The simulations agree well with the experimental results for retention time, and they can also reasonably capture the peak shapes in all four different gradients. Furthermore, the model can well describe the relative retention of different proteins at a given gradient. The only exception is GUSD15, whose experimental peak width is broader. Moreover, the fusion protein elution under different gradient slopes follows the typical trends (Gallant, et al., 1996a, 1996b; Luo and Hsu, 1997). When the gradient slope becomes steeper, the protein peak moves towards higher salt concentration and becomes sharper.

Prediction of protein elution

Based on the protein parameters, Z and I, Eq. (13) was used to calculate the salt concentration at which a protein's peak is eluted during linear gradient elution. The calculated values for each GUS fusion under different gradient slope as well as the protein retention from the simulation are compared with the experimental results in Figure 4.7. In most of cases, both models predict the protein peak elution with an error of less than 10%. The likely reason for the larger errors seen for GUSD0 is that the presence of an overlapping E. coli peak made determination of isocratic elution times (and, therefore, Z and I) more difficult. As expected, the predicted retentions for the two models are very similar because both assume instantaneous equilibrium between the mobile and stationary phase.

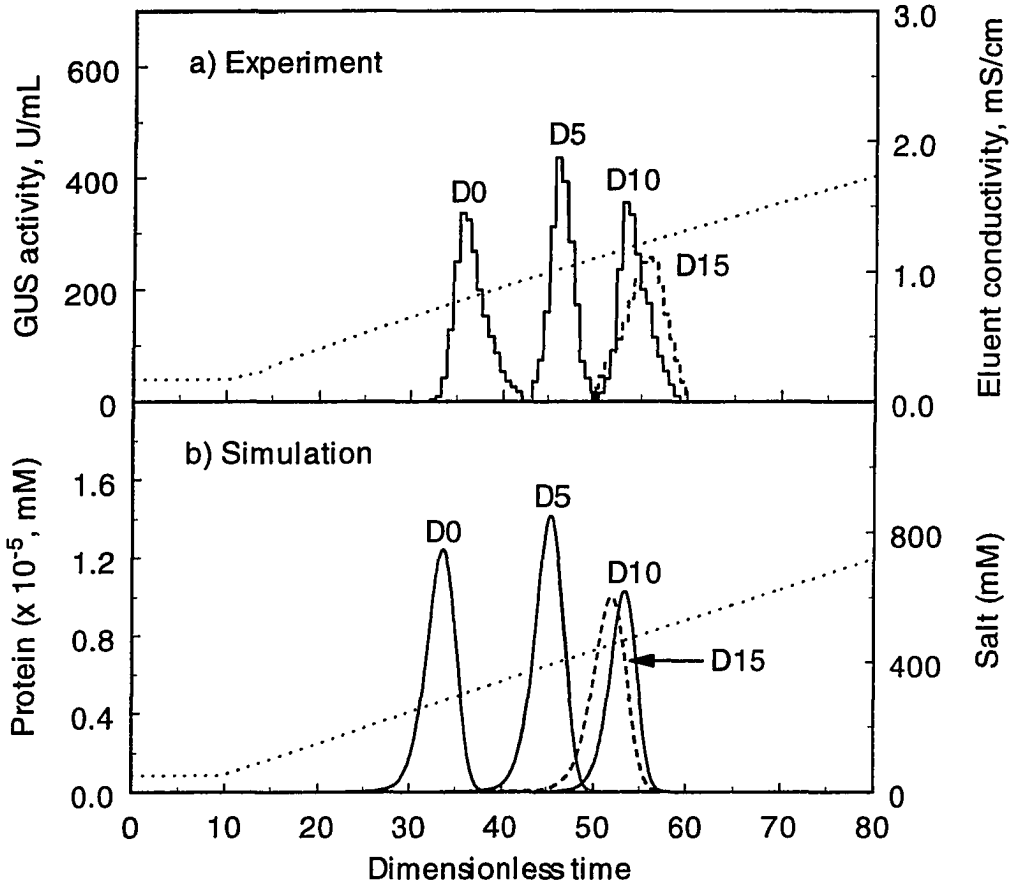


Figure 4.3. Comparison of experimental and simulated chromatograms. a) Reconstructed GUS activity peaks under linear gradient elution. Feed injection: 250 μl of samples (0.053 column dead volumes) with protein concentrations listed in Table 4.3. Operating conditions: 1 ml/min, pH 7.0, 3 ml fractions. b) Simulated chromatograms. Gradient slope: 9.5 mM chloride ion per column dead volume.

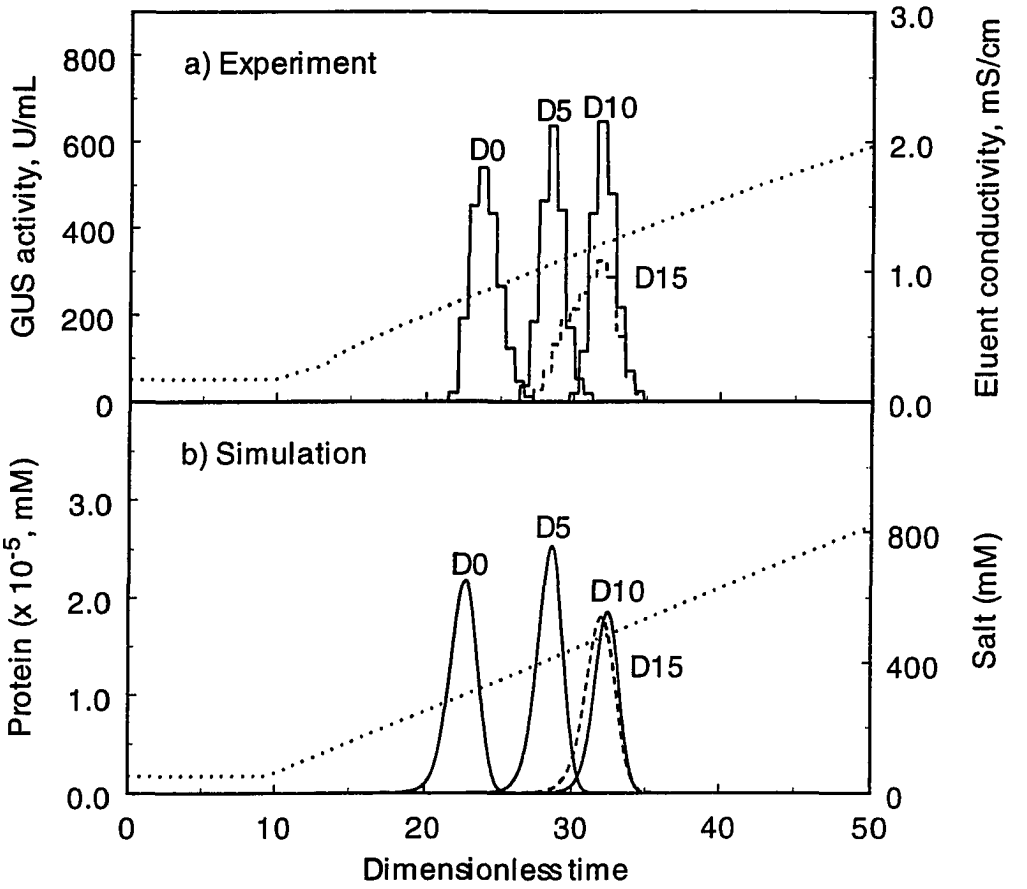


Figure 4.4. Comparison of experimental and simulated chromatograms. a) Reconstructed GUS activity peaks. b) Simulated chromatograms. Gradient slope: 19 mM chloride ion per column dead volume.

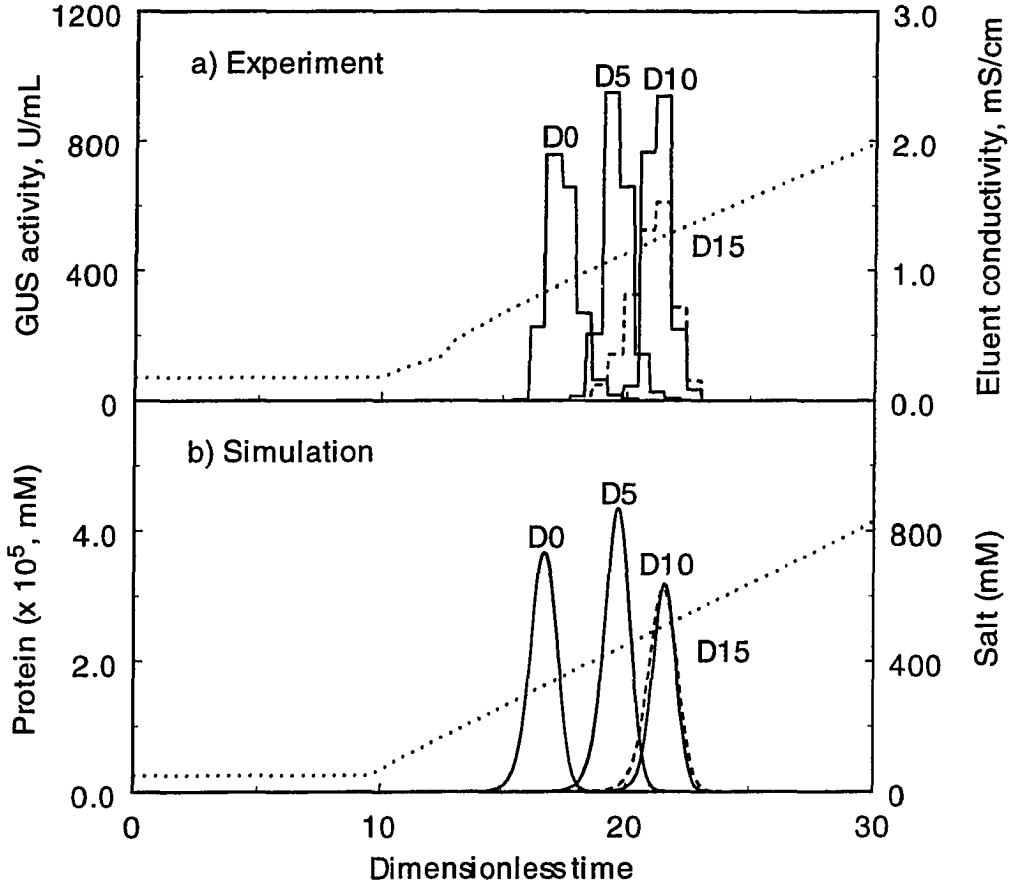


Figure 4.5. Comparison of experimental and simulated chromatograms. a) Reconstructed GUS activity peaks. b) Simulated chromatograms. Gradient slope: 38 mM chloride ion per column dead volume.

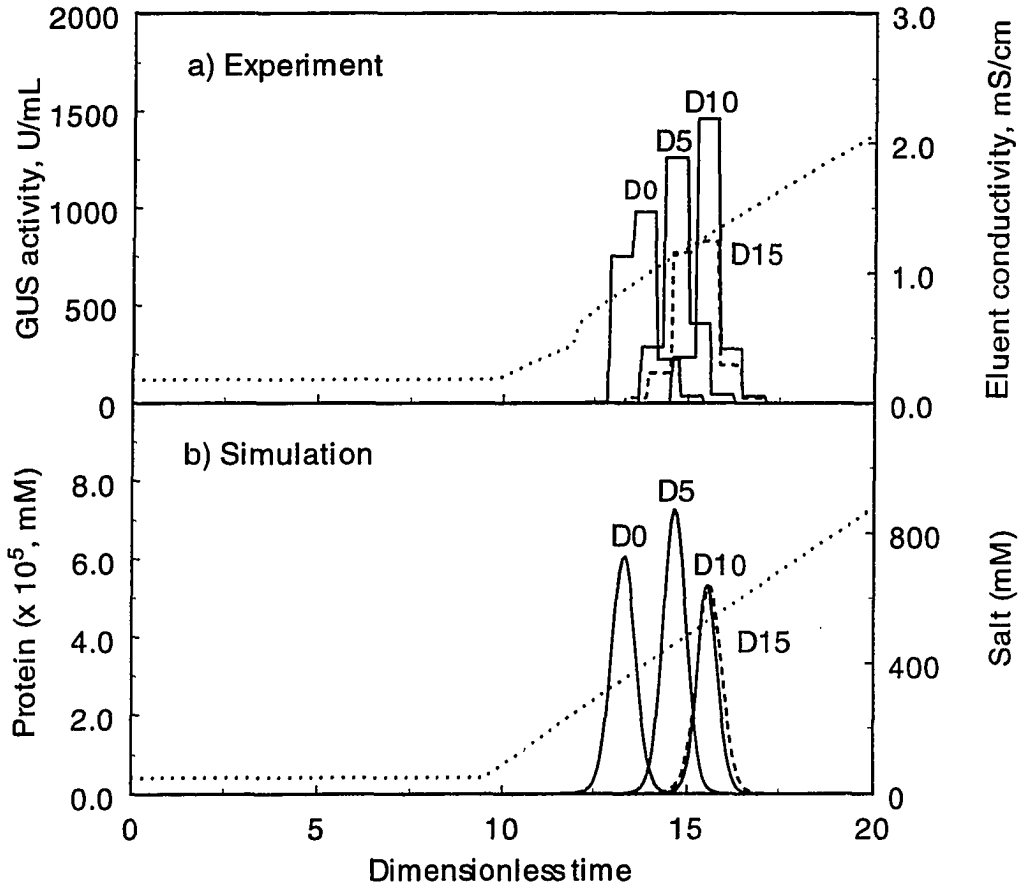


Figure 4.6. Comparison of experimental and simulated chromatograms. a) Reconstructed GUS activity peaks. b) Simulated chromatograms. Gradient slope: 79.2 mM chloride ion per column dead volume.

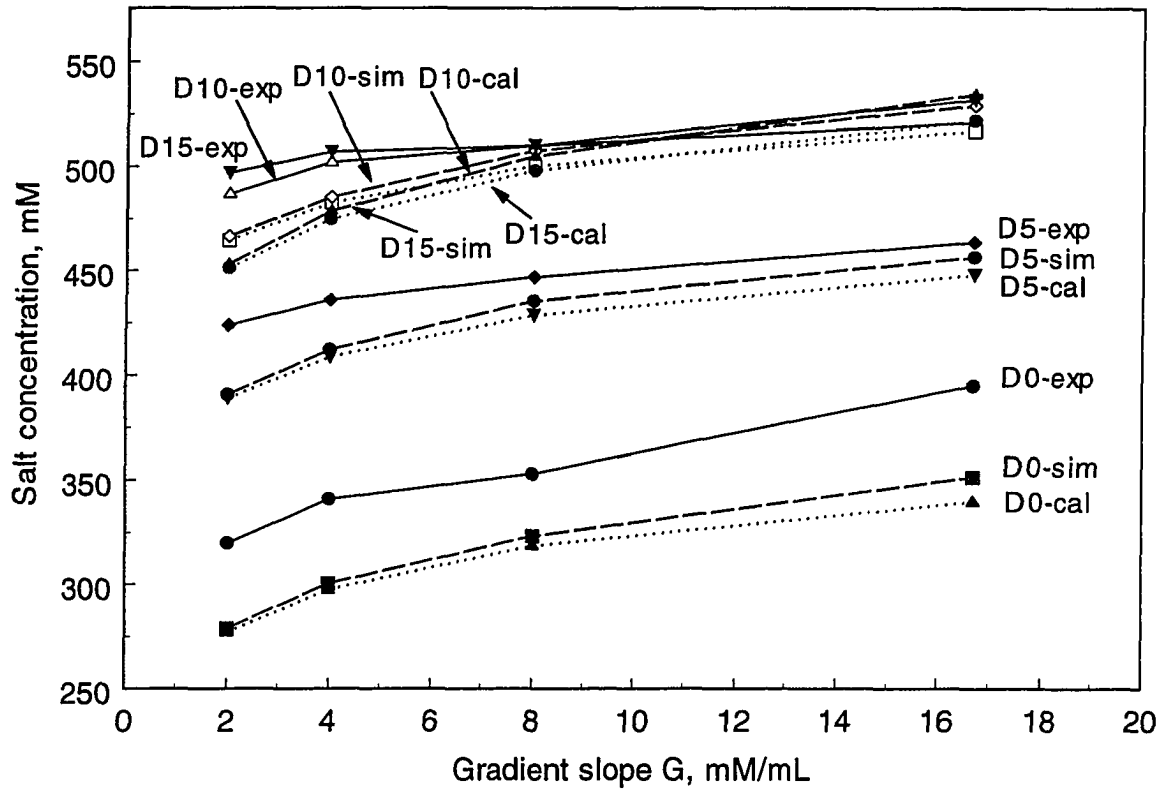


Figure 4.7. Comparison of predicted protein retention (salt concentration at elution) by the lumped dispersion model (simulation) and Equation 13 (calculation) with experimental results under different gradient slope (G). Cal = calculation; Sim = simulation; Exp = experiment.

Conclusions

Fusion proteins are characterized by Z and I based on the stoichiometric displacement model under isocratic elution conditions. When the fusion tail is long enough (GUSD15) the protein shows a different change trend in Z and I values. With addition of 5 aspartates in the tail from GUSD0 to GUSD5 to GUSD10, the characteristic charge of protein increases by 4, and the equilibrium constant increases by 220 times. From GUSD10 to GUSD15, the addition of 5 aspartates results in decreased characteristic charge and protein equilibrium constant.

Protein specific parameters obtained from isocratic elution are used in two models to predict the protein retention in gradient elution. It is shown that both models can well describe the fusion protein retention in gradient elution ion-exchange process at different gradient slopes, and the numerical simulation with axial dispersion effect included can additionally capture the protein elution peak shapes.

Acknowledgment

This work was supported by the NSF under Grant No. BES9522644. The authors thank ProdiGene Inc., College Station, TX, for providing us *E. coli* strains encoded with GUS and its fusions.

List of Symbols

C_b concentration of the salt ion in stationary phase, (mM)

C_0 concentration of the salt ion in mobile phase, (mM)

$C_{0,e}$	counter ion concentration of the equilibrating buffer, (mM)
D_i	effective dispersion coefficient, (cm^2/s)
F	volumetric flow rate (ml/s)
G	gradient slope in dimensionless time scale, mM
G'	gradient slope in mobile phase volume scale (mM/mL)
H	height equivalent to a theoretical plate, (m)
I	solute specific constant
k'	capacity factor
K_{eq}	equilibrium constant
L	column length (m)
P_0	protein concentration in the mobile phase, (mM)
P_b	protein concentration on the stationary phase, (mM)
$P_{0,f}$	protein concentration in a sample (mM)
t	time dimension (s)
t_0	column dead time, retention time of solute at non-retained condition (s)
t_R	retention time of a solute at retained condition (s)
u_0	the chromatographic velocity ($u_0 = u_s / \varepsilon$) (cm/s)
u_l	linear velocity, (cm/s)
V_d	dead volume of connection tube between the outlet of the gradient generating device and the top of the column (ml)
V_m	total volume of the mobile phase in the column (ml)

X	axial dimension of column (cm)
x	nondimensional axial position ($x = X / L$)
Z	the number of bound ions displaced during the protein adsorption process
ε	column total porosity
τ	nondimensional time ($\tau = t / t_0$)
τ_f	length of the sample feed pulse
φ	phase ratio (ratio of the stationary and mobile phase volume in a column)
Λ	ion capacity of the resin (mM)
χ	gradient shape, linear gradient $\chi = 1$

Literature Cited

Berninger, J. A.; Whitley, R. D.; Zhang, X.; Wang, N.-H. L. A versatile model for simulation of reaction and nonequilibrium dynamics in multicomponent fixed-bed adsorption processes. Comput. Chem. Eng. 1991, 15, 749-768.

Boardman N. K. and Partridge, S. M. Separation of neutral proteins on ion-exchange resins. Biochem. J. 1955, 59, 543-552.

Bonnerjea, J.; Oh, S.; Hoare, M.; Dunnill, P. Protein purification: the right step at the right time. Bio/Technol. 1986, 4, 954-958.

Brooks, C. A.; Cramer, S. M. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. AIChE J. 1992, 38, 1969-1978.

Chicz, R. M.; Regnier, F. E. Single amino acid contributions to protein retention in cation-exchange chromatography: resolution of genetically engineered subtlisin variants. Anal. Chem. 1989, 61, 2059-2066.

Czok, M.; Guiochon, G. The physical sense of simulation models of liquid chromatography: propagation through a grid or solution of the mass balance equation. Anal. Chem. 1990, 62, 189-200.

Drager, R. R.; Regnier, F. E. Application of the stoichiometric displacement model of retention to anion-exchange chromatography of nucleic acids. J. Chromatogr. 1986, 359, 147-155.

Drager, R. R.; Regnier, F. E. Retention mechanism of lactate dehydrogenase in anion-exchange chromatography. J. Chromatogr. 1987, 406, 237-246.

Gadam, S. D.; Jayaraman, G.; Cramer, S. M. Characterization of non-linear adsorption properties of dextran-based polyelectrolyte displacers in ion-exchange systems. J. Chromatogr. 1993, 630, 37-52.

Gallant, S. R.; Vunnum, S.; Cramer, S. M. Optimization of preparative ion-exchange chromatography of proteins: linear gradient separations. J. Chromatogr. A 1996a, 725, 295-314.

Gallant, S. R.; Vunnum, S.; Cramer, S. M. Modeling gradient elution of proteins in ion-exchange chromatography. AIChE J. 1996b, 42, 2511-2520.

Gerstner, J. A.; Bell, J. A.; Cramer, S. M. Gibbs free energy of adsorption for biomolecules in ion-exchange systems. Biophys. Chem. 1994, 52, 97-106.

Heng, M. H.; Glatz, C. E. Charged fusions for β-galactosidase retention in anion-exchange chromatography. J. Chromatogr. A 1995, 689, 227-234.

Jandera, P.; Churacek, J. Gradient elution in liquid chromatography I. The influence of the composition of the mobile phase on the capacity ratio (retention volume, band width, and resolution) in isocratic elution - theoretical considerations. J. Chromatogr. 1974a, 91, 207-221.

Jandera, P.; Churacek, J. Gradient elution in liquid chromatography II. Retention characteristics (retention volume, band width, resolution, plate number) in solvent-programmed chromatography - theoretical considerations. J. Chromatogr. 1974b, 91, 223-235.

Jefferson, R. A.; Burgress, S. M.; Hirsh, D. β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 1986, 83, 8447-8451.

Jefferson, R. A.; Wilson, K. J. The GUS gene fusion system. Plant Mol. Biol. Manual 1991, B14, 1-33.

Kato, Y.; Komiya, K. Hishimoto, T. Study of experimental conditions in high-performance ion-exchange chromatography of proteins. J. Chromatogr. 1982, 246, 13-22.

Kato, K.; Yoshida, K.; Tsukamoto, H.; Nobunaga, M.; Masuya, T.; Sawada, T. Synthesis of *p*-nitrophenyl β-D-glucosiduronic acid and its utilization as a substrate for the assay of β-Glucuronidase activity. Chem. Pharm. Bull. 1960, 8, 239-242.

Kawasaki, T. and Bernardi, G. Investigations on the resolving power of hydroxyapatite columns. Biopolymers, 1970, 9, 257-268.

Kim, S. U.; Berninger, J. A; Yu, Q.; Wang, N.-H. L. Peak compression in stepwise pH elution with flow reversal in ion exchange chromatography. Ind. Eng. Chem. Res. 1992, 31, 1717-1730.

Kopaciewicz, W.; Rounds, M. A.; Fausnaugh, J.; Regnier, F. E. Retention model for high-performance ion-exchange chromatography. J. Chromatogr. 1983, 266, 3-21.

Luo, R. G.; Hsu, J. T. Optimization of gradient profiles in ion-exchange chromatography for protein purification. Ind. Eng. Chem. Res. 1997, 36, 444-450.

Ma, Z.; Guiochon, G. Application of orthogonal collocation on finite elements in the simulation of non-linear chromatography. Comput. Chem. Eng. 1991, 15, 415-426.

Morris, C. J. O. R. and Morris, P. Separation methods in biochemistry. 2nd ed. Pitman Publishing, 1976.

Niederauer, M. Q.; Suominen, I.; Rougvie, M. A.; Ford, C. F.; Glatz, C. E. Characterization and polyelectrolyte precipitation of β-galactosidase containing genetic fusions of charged polypeptides. Biotechnol. Prog. 1994, 10, 237-245.

Novotny, J. Chromatography of proteins and peptides on sephadex ion-exchangers: dependence of the resolution on the elution schedule. FEBS Lett. 1971, 14, 7-10.

Regnier, F. E. and Mazsaroff, I. A theoretical examination of adsorption processes in preparative liquid chromatography of proteins. Biotechnol. Prog. 1987, 3, 22-26.

Rhee, H.-K. and Amundson, N. R. A study of the shock layer in nonequilibrium exchange system. Chem. Eng. Sci. 1972, 27, 199-211.

Rounds, M. A.; Regnier, F. E. Evaluation of a retention model for high-performance ion-exchange chromatography using two displacing salts. J. Chromatogr. 1984, 283, 37-45.

Sassenfeld, H. M.; Brewer, S. J. A polypeptide fusion designed for the purification of recombinant proteins. Bio/Technol. 1984, 2, 76-81.

Snyder, L. R. Linear elution adsorption chromatography VII. Gradient elution theory. J. Chromatogr. 1964, 13, 415-434.

Snyder, L. R.; Dolan, J. W.; Gant, J. R. Gradient elution in high-performance liquid chromatography. I. Theoretical basis for reversed-phase systems. J. Chromatogr. 1979, 165, 3-30.

Stout, R. W.; Sivakoff, S. I.; Ricker, R. D.; Snyder, L. R. Separation of proteins by gradient elution from ion-exchange columns: optimizing experimental conditions. J. Chromatogr. 1986, 353, 439-463.

Velayudhan, A. and Horvath, C. Preparative chromatography of proteins: Analysis of the multivalent ion-exchange formalism. J. Chromatogr. 1988, 443, 13-29.

Velayudhan, A. and Horvath, C. Adsorption and ion-exchange isotherms in preparative chromatography. J. Chromatogr. 1994, 663, 1-10.

Yamamoto, S.; Nakanishi, K.; Matsuno, R.; Kamikubo, T.; Ion exchange chromatography of proteins -- prediction of elution curves and operating conditions. I. Theoretical considerations. Biotechnol. Bioeng. 1983a, 25, 1465-1483.

Yamamoto, S.; Nakanishi, K.; Matsuno, R.; Kamikubo, T.; Ion exchange chromatography of proteins -- prediction of elution curves and operating conditions. II. Experimental verification. Biotechnol. Bioeng. 1983b, 25, 1373-1391.

Yamamoto, S.; Nomura, M.; Sano, Y. Adsorption chromatography of protein determination of optimum conditions. AIChE J. 1987, 33, 1425-1434.

Yamamoto, S.; Nakanishi, K.; Matsuno, R. Ion-exchange chromatography of proteins. Marcel Dekker, Inc., New York, 1988.

Zhang, C.-M.; Glatz, C. E. Processing engineering strategy for recombinant protein recovery from canola by cation exchange chromatography. Biotechnol. Prog. 1999, 15, 12-18.

Zhang, C.-M.; Love, R. T.; Jilka, J. M.; Glatz, C. E. Genetic engineering strategies for purification of recombinant proteins from canola by anion exchange chromatography: An example of β-glucuronidase. 1999, submitted to Biotechnol. Prog.

CHAPTER 5. SUITABILITY OF IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY FOR PROTEIN PURIFICATION FROM CANOLA

A paper to be submitted to Biotechnology and Bioengineering

Chen-Ming Zhang[1], Susan A. Reslewic[2], and Charles E. Glatz[3]

Abstract

This work demonstrated that proper selection of a metal ion and chelating ligand enables recovery of a his_6-tagged protein from canola extracts by immobilized metal affinity chromatography (IMAC). When using Co^{2+} with iminodiacetate (IDA) as the chelating ligand, β-glucuronidase-his_6 (GUSH6) can be purified from canola protein extract with almost homogeneous purity in a single chromatographic step. The discrimination with which metal ions bound native canola proteins followed the order $Cu^{2+} < Ni^{2+} < Zn^{2+} < Co^{2+}$ as regards elimination of proteins co-eluted with the fusion protein. IDA and nitrilotriacetate (NTA) immobilized metal ions showed different binding patterns, whose cause is attributed to a more flexible coordination binding that allows for multi-site interactions.

---

[1] Graduate student, primary researcher and author

[2] Did some preliminary work on GUS-His_6 spiked canola extract on Ni-IDA column

[3] Adviser of the primary author and author for correspondence

Introduction

Since its introduction in 1975 (Porath, et al., 1975), immobilized metal affinity chromatography (IMAC) has quickly become one of the separation techniques of choice for purification of many proteins and peptides in both analytical and large-scale modes (Porath, 1988a; Hansen, et al., 1992; Lindner, et al., 1992; Ritter, et al., 1992; Sharma, et al., 1992; Beitle and Ataai, 1993; Wilkinson, et al., 1995; Oswald, et al., 1997; see Wong, et al., 1991 for a more detailed list). IMAC separation exploits the affinities between the immobilized metal ions and electron donating groups on protein surface. The resin is stable and inexpensive, and in some instances, provides similar selectivity as "biospecific" affinity separations (Arnold, 1991; Porath, 1988b). The dominating electron donating group in a protein is the imidazole side chain of histidine, while the N-terminus of the protein contributes to a lesser extent. In addition, the thiol group of cysteine would be a good electron donor, but it is rarely present in the appropriate, reduced state (Arnold, 1991; Sulkowski, 1985).

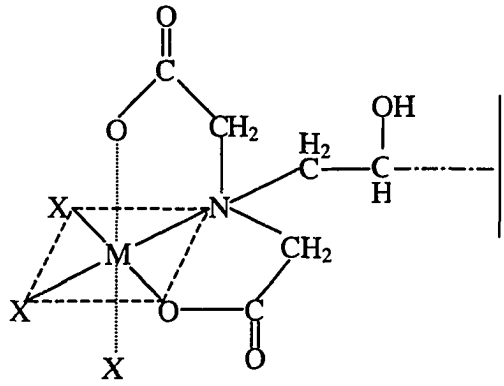
The favored genetic engineering method for enhancing a protein's affinity to metal sites has been to attach a six-histidine tail (his_6) on either the C- or N-terminus (Oswald, et al., 1997; Wilkinson, et al., 1995; Lindner, et al., 1992; Hochuli and Piesecki, 1992; Hochuli, et al., 1988). With an added tail, an otherwise non-binding protein from a cell lysate can be purified to 90% or even better purity in a single chromatographic step (Oswald, et al., 1997; Wilkinson, et al., 1995; Lindner, et al., 1992; Hochuli and Piesecki, 1992; Hochuli, et al., 1988).

The most widely used metal ions in IMAC are the first row transition metals, Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+} (Porath, 1988b; Sulkowski, 1989; Wong, et al., 1991), though other

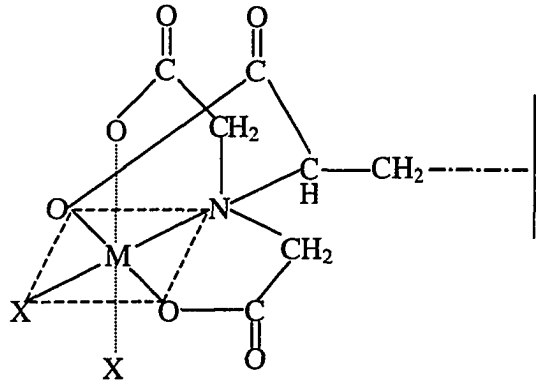
metal ions have been tried (Porath, et al., 1983; Mantovaara, et al., 1989; Andersson, 1991). Each of the four metal ions can coordinate up to six electron donors, and their affinity towards imidazole follows the order $Cu^{2+}>Ni^{2+}>Zn^{2+}\sim Co^{2+}$ (Sundberg and Martin, 1974). The metal ions can be immobilized by various chelating ligands including, iminodiacetate (IDA), tris (carboxymethyl) ethylene-diamine (TED) (Sulkowski, 1985), nitrilotriacetate (NTA) (Hochuli, et al., 1987), carboxymethylated aspartate (CM-ASP), tetraethylene pentamine (TEPA) (Wong, et al., 1991). Among these, only IDA and NTA-derived resins are commercially available. IDA forms tridentates with metal ions (Sulkowski, 1985), while NTA forms tetradentates (Hochuli, et al., 1987) (Fig. 5.1). This leaves three free orbitals for IDA-Me and two for NTA-Me to interact with protein molecules for metal ions of coordination number six.

Although IMAC was introduced more than two decades ago, its application has been mainly to purification of proteins from microbial production systems. As transgenic plants emerge as alternative hosts for recombinant proteins, a systematic study of their suitability for use with IMAC is merited.

Canola, the plant material used here, has been studied for its suitability as a host for recombinant protein production from the protein purification point of view (Zhang and Glatz, 1999; Zhang, et al., 1999). It is one of the objectives of this work to study the feasibility and effectiveness of purifying a his_6-tagged protein, β-glucuronidase-his_6 (GUSH6), from canola extract by a single IMAC step. The GUSH6 binding behaviors on IDA and NTA immobilized metal ions are compared, and causes for the differences considered.



(a) IDA-Me^{2+} chelate



(b) NTA-Me^{2+} chelate

Figure 5.1. Schematic structures of different chelate ligands immobilized metal ions having a coordination number of six. (a) Iminodiacetate (IDA) immobilizes a metal ion by forming a tridentate, and this leaves three free orbitals on the metal ion to interact with proteins. (Sulkowski, 1985). (b) Nitrilotriacetate (NTA) immobilizes a metal ion by forming a tetradentate, and this leaves two free orbitals on the metal ion to interact with proteins (Hochuli, et al., 1987).

Materials and Methods

GUSH6

E. coli GUS is a tetramer with monomeric molecular weight of ~68000 (Jefferson et al., 1986). *E. coli* strains encoded with the GUS gene carrying a six histidine C-terminal fusion tag were obtained from ProdiGene (College Station, TX) and cultured in LBH media, where GUS expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG). Cells harvested by centrifugation were resuspended in buffer containing 50 mM NaPi and 1 MNaCl (pH 7.0). Phenylmethylsulfonyl fluoride (PMSF, 2 mM final concentration) was added to inhibit *E. coli* proteases (Jefferson and Wilson, 1991) before sonication (5 × 30 seconds pulse) to disrupt the cells. After centrifugation (15,000 rpm, 20 min), the supernatant was collected for recovery of the GUSH6.

The supernatant was loaded to an IMAC column (Ni^{2+}-IDA on crosslinked agarose, Novagen, Madison, WI) equilibrated in 50 mM NaPi, 1M NaCl, pH 7.0 (buffer A). Elution was with an increasing linear gradient of buffer B (1M imidazole, 50 mM NaPi, 1 M NaCl, pH 7.0) from 0 to 50% B. The fractions with GUS activity were concentrated and buffer-exchanged to buffer A by diafiltration using PM30 membrane (MWCO=30K, Amicon, Beverly, MA). The final GUSH6 stock solution had a concentration of 10,560 U/ml and a specific activity of 32,000 U/mg.

Protein assays

The GUS activity assays were carried out following the method described in Zhang, et al. (1999). The substrate for the enzymatic reaction was *p*-nitrophenyl β-D-glucosiduronic acid (PNPG) (Kato et al., 1960; Jefferson and Wilson, 1991). One unit of GUS denotes liberation of 1 nmol *p*-nitrophenol/min from PNPG at 37 °C and pH 7.0. The protein mass

was determined by the Bio-Rad protein assay (Bio-Rad, Hercules, CA), with bovine serum albumin (BSA) as the standard.

Canola protein extract

Non-transgenic canola seed was provided by Pioneer Hi-Bred International (Johnston, IA). The canola seed was processed by grinding, dehulling, flaking, and extraction of oil with *n*-hexane extraction (1 g meal/20 mL of hexane) in two stages, 30 minute each, at room temperature to a final oil content <2% (w/w) (Zhang and Glatz, 1999).

Canola extract was obtained by mixing defatted canola meal with chromatography buffer A at 1:10 (w/v) for 30 min (Zhang and Glatz, 1999). The extract had a protein concentration of 23.3 ± 0.2 mg/mL. By comparison, extraction with the buffer used previously (50 mM NaPi pH 7.0) gave extracts with a protein concentration of 21.3 ± 0.2 mg/ml. Both buffers extracted equal amounts of GUS from transgenic canola.

Immobilized metal affinity chromatography

Chromatographic experiments were carried out using a fast performance liquid chromatography (FPLC) system controlled by BioLogic software (Bio-Rad, Hercules, CA). Resins were degassed before packing into 15 × 10 Econo columns (Bio-Rad, Hercules, CA) to a total bed volume of 5 mL. A flow adaptor, which was inserted onto the top of the settled bed, was used to prevent sample application and gradient elution from dilution and ensure even sample distribution in the column. Packed columns were first stripped (resins came with immobilized Ni^{2+}) by 50 mM EDTA in 50 mM NaPi, 1M NaCl, pH 7.0 solution (5 CV) before charging with selected metal ions. The metal ion solutions used to charge the column were 50 mM in the metal using: $NiSO_4$, pH 5.35; $ZnSO_4$, pH 5.16; $CoSO_4$, pH 5.38; and $CuSO_4$, pH 4.58. Equilibrating (buffer A) and elution buffers (buffer B) were the same as

described previously. To assure consistent experimental conditions, the column regeneration procedures listed in Table 5.1 was followed before each run. Column flow rate was 1 mL/min.

"Spiked" chromatographic samples were prepared by adding GUSH6 stock solution into canola extract (2:5, v/v) to simulate transgenic canola extract (for complete experimental flow chart, see Zhang and Glatz, 1999). Sample was loaded to a 1 mL sample loop and injected into the column. Imidazole gradient (0 to 500 mM imidazole over a 100 mL (20 CV) elution volume) was applied for bound protein elution based on the preliminary work on Ni^{2+}-IDA showing that a pH gradient from 7.9 to 4.0 reduced GUS recovery by ~60%, and steeper imidazole gradients (0 - 1000 mM, 12 CV) reduced resolution between bound canola proteins

Table 5.1. IMAC column regeneration protocol. The flow rate for all the procedures below was 1.2 mL/min. All solutions used were degassed.

Column volumes	Solution
5	50 mM EDTA in buffer A*, pH 7.0
5	0.2 M NaOH
5	Deionized water
5	50 mM metal ion solution
5	0.1 M NaAc, 1 M NaCl, pH 7.0
3	Deionized water
various	Buffer A

* Buffer A = 50 mM NaPi, 1 M NaCl, pH 7.0.

and GUSH6. Fractions were collected and assayed for protein concentration and GUS activity.

Materials

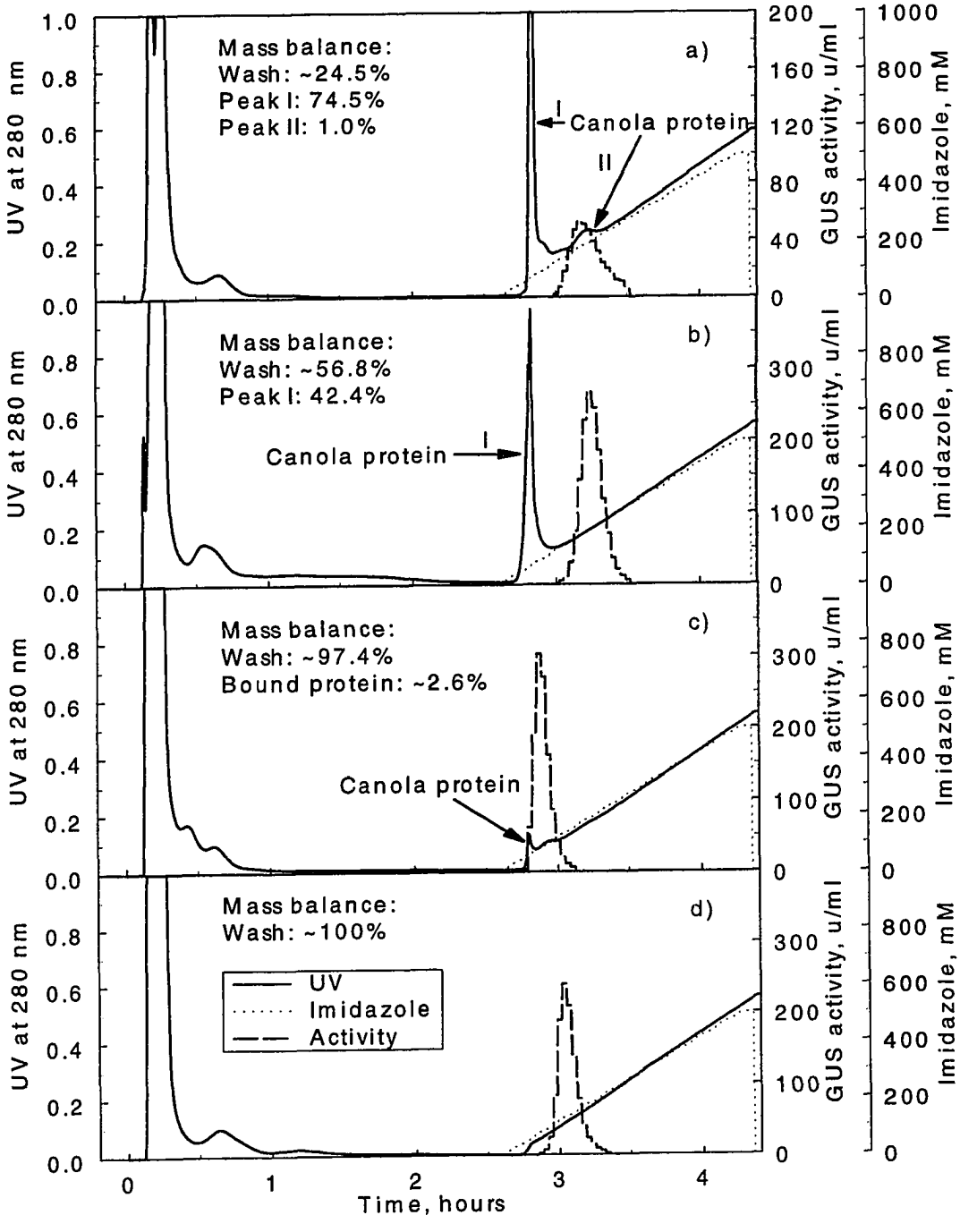
Ni-IDA resin (minimum Ni^{2+} loading capacity 20 μmol/ml resin) was purchased from Novagen (Madison, WI). Ni-NTA superflow resin (Ni^{2+} loading capacity 17 μmol/ml resin) was purchased from Qiagen (Valencia, CA). PNPG and BSA standard were purchased from Sigma (St. Louis, MO). All other chemicals were purchased from Fisher (Itasca, IL).

Results and Discussion

GUSH6 purification from canola extract by IDA immobilized metal ions

Spiked canola samples were run on the same column charged with different metal ions: Cu^{2+}, Ni^{2+}, Zn^{2+}, and Co^{2+}. The chromatograms obtained from each column are shown in Figure 5.2, and the relative retentions of GUSH6 on different Me^{2+}-IDA columns are shown in Figure 5.3. Figure 5.2 shows that different metal ions have different degrees of binding to GUSH6 and native canola proteins. All the chromatograms show large wash peaks, which include absorbance from both unbound protein and canola pigments. From Fig. 5.2, the amount of canola protein binding on a Me^{2+}-IDA column decreases following the order Cu^{2+}(75.5%), Ni^{2+}(42.4%), Zn^{2+} (2.6%), and Co^{2+} (~0%). This trend agrees well with the reported recognition tendencies of IDA-immobilized metal ions for various surface histidine residue distributions (Table 5.2). According to Table 5.2, a protein needs at least two histidines to be retained on a Ni^{2+}/Zn^{2+}/Co^{2+}-IDA column, whereas Cu^{2+}-IDA can recognize any histidine distribution, even single histidines, on a protein surface. From this

Figure 5.2. GUSH6 spiked canola extract on different metal ion charged IDA-IMAC columns. a) Cu^{2+}. b) Ni^{2+}. c) Zn^{2+}. d) Co^{2+}. Flow rate = 1 mL/min. Sample volume = 1 mL. Gradient slope: 5 mM/mL. Fraction size = 2 mL. Mass balances for wash peaks were obtained by subtracting the bound protein (eluted after gradient elution started) from the total loaded protein.


Mass balance:
Wash: ~24.5%
Peak I: 74.5%
Peak II: 1.0%
a)
I
II
Canola protein
Mass balance:
Wash: ~56.8%
Peak I: 42.4%
b)
Canola protein
I
Mass balance:
Wash: ~97.4%
Bound protein: ~2.6%
c)
Canola protein
Mass balance:
Wash: ~100%
d)
UV
Imidazole
Activity
UV at 280 nm
GUS activity, u/ml
Imidazole, mM
Time, hours

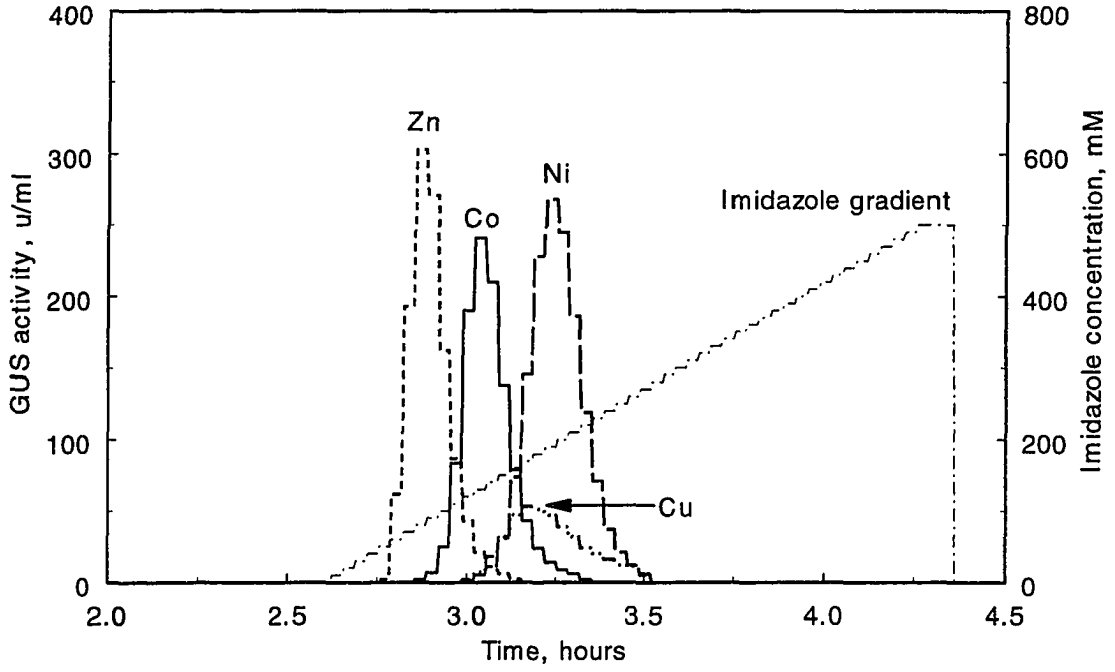


Figure 5.3. Relative retentions of GUSH6 on different Me-IDA columns. All experiments started with approximately the same total GUS activity. GUSH6 has a very low recovery on Cu^{2+}-IDA.

Table 5.2. Recognition of Me^{2+}-IDA on different surface topography of histidine residues. Buffer: 20 mM PB, 1 M NaCl, pH 7.0 (Sulkowski, 1989). (+) = retention; (-) = no observed retention.

Ligand	Cu	Ni	Zn	Co
-His-	+	-	-	-
$-His(X)_nHis-$	+	+	-	-
$-His(X)_nHis-$ n(2,3); α-Helix	+	+	+	+
-HisHis-	+	+	+	+

we would estimate that most canola proteins have at least one accessible histidine, and less than 50% of canola proteins have accessible, multiple histidine sites.

The binding behaviors of native canola proteins on Zn^{2+}-IDA and Co^{2+}-IDA were expected to be very similar according to Table 5.2. Figure 5.2 shows that the difference, although small, is noticeable (Figures 5.2c and d). While 2.6% of native canola protein bound to Zn^{2+}-IDA, almost no protein other than GUSH6 bound to Co^{2+}-IDA. Hence, the recognition ability (RA) of Me^{2+}-IDA on protein surface histidines follows the order $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$. For differentiation ability the reversed order holds.

IDA-IMAC was able to purify the six-histidine fusion protein from canola extract. The GUSH6 recoveries and enrichment ratios after the single chromatographic step are listed in Table 5.3. Even though Zn^{2+}-IDA bound less canola protein than Ni^{2+}-IDA, the GUSH6 peak in Zn^{2+}-IDA overlapped with bound canola proteins (Figures 5.2b and c) and resulted in a similar GUSH6 enrichment ratio as on Ni^{2+}-IDA. The Co^{2+}-IDA column was able to purify

Table 5.3. GUSH6 recoveries from canola extract by a single chromatographic run on various Me^{2+}-IDA columns.

Metal ions	Cu^{2+}	Ni^{2+}	Zn^{2+}	Co^{2+}
Activity before loading, U/mg	190	190	180	200
Final specific activity, U/mg	5,000	20,000	17,000	141,000
Recovery, %	25.5	85.8	71.8	66.8
Enrichment ratio*	26	105	94	705

* Enrichment ratio = Final specific activity/specific activity before loading.

GUSH6 to almost homogeneous purity. SDS-PAGE (not shown here) of fractions (25 times concentrated) with the six highest GUS activities collected in Co^{2+}-IDA experiment showed no visible native canola protein bands.

The recovery of GUSH6 on Cu^{2+}-IDA was significantly lower than on other Me^{2+}-IDA (Table 5.3). This phenomenon can probably be attributed to one or more of the following three factors. The first is that immobilized Cu^{2+} is not entirely redox-stable and could be involved in catalytic oxidation reactions of GUSH6 deactivation (Winzerling, et al., 1992). The second is the possibility of metal ion leakage during imidazole gradient elution. The last is the metal ion transfer (MIT) process (Sulkowski, 1989), that is the binding of GUSH6 with Cu^{2+} is so strong that the Cu^{2+}-IDA bonds are broken during the imidazole elution process. The result is the elution of inactive GUSH6. That free Cu^{2+} does inactivate GUSH6 is seen in Figure 5.4 where there is a sudden loss and continuing decline of GUS activity once Cu^{2+} is introduced into solutions containing GUSH6. The presence of imidazole slows the inactivation.

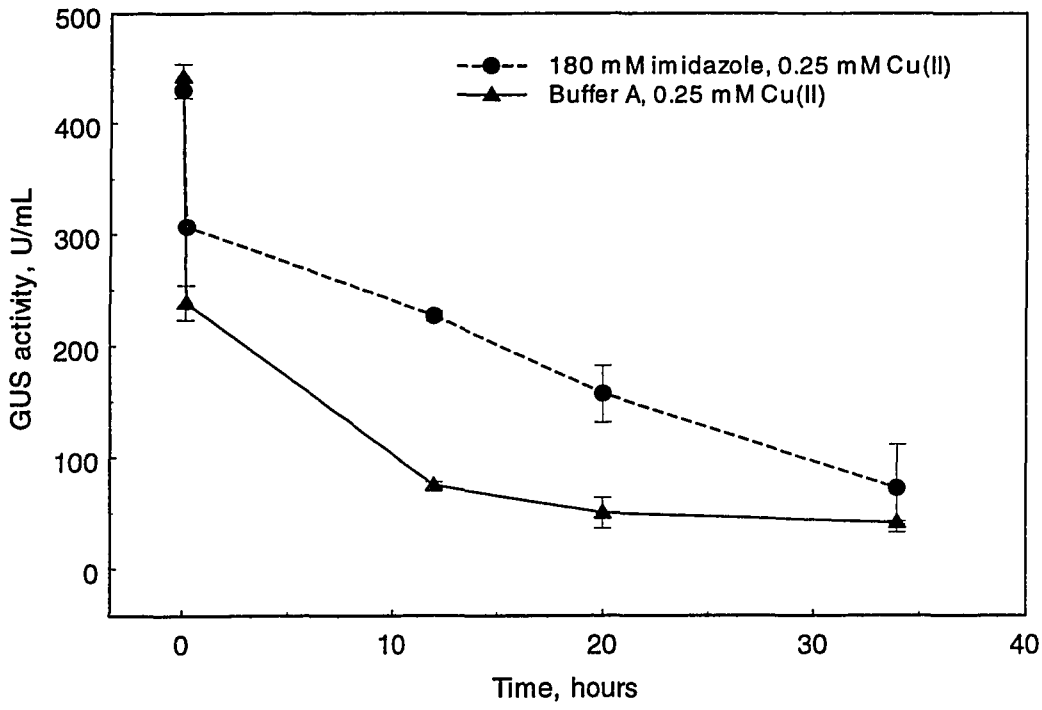


Figure 5.4. The influence of free Cu^{2+} ion on GUS activity. Buffer A = 50 mM NaPi, 1 M NaCl, pH 7.0. 180 mM imidazole = buffer A with 180 mM imidazole (the imidazole concentration eluting GUSH6 on Cu^{2+}-IDA). GUSH6 was first mixed with selected buffer solution to provide the initial GUS activities, then $CuSO_4$ solution was added. The solutions were held at 4°C from 0 to 12 hours, and at room temperature afterwards.

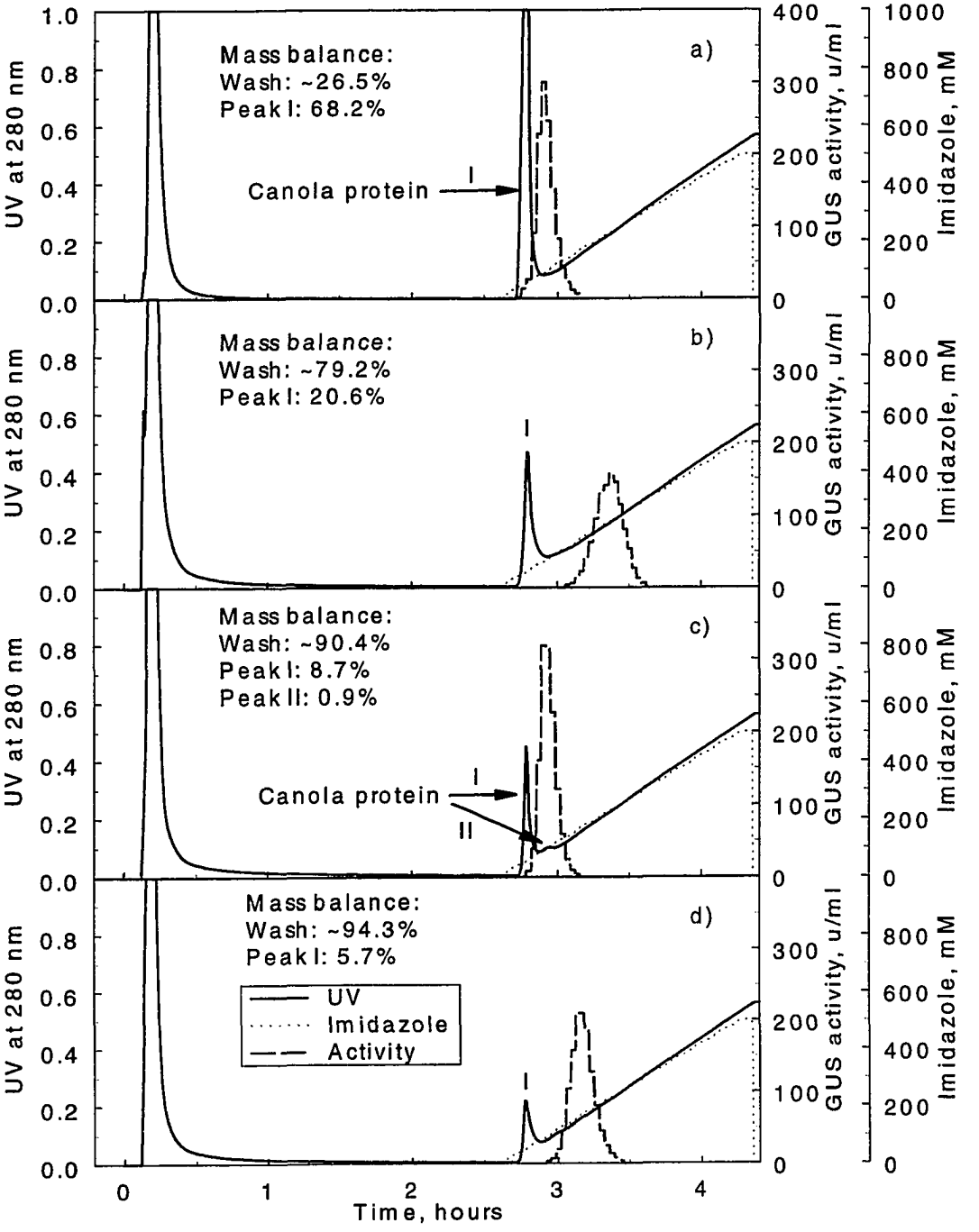
A comparison of Figure 5.3 and Table 5.3 shows that binding affinity (as measured by the imidazole concentration needed for protein elution) does not translate directly to purification performance. A higher imidazole concentration is needed to elute GUSH6 on Ni^{2+}-IDA than Co^{2+}-IDA (and also for Cu^{2+}-IDA relative to Zn^{2+}-IDA), but higher purification results from the earlier elution. Nor can the differentiation pattern of Table 5.2 be a conclusive guide as the comparison of Ni^{2+}-IDA and Zn^{2+}-IDA performance shows.

GUSH6 purification from canola extract by NTA immobilized metal ions

GUSH6 spiked canola samples were also run on columns with Ni^{2+}, Zn^{2+}, Co^{2+}, and Cu^{2+} immobilized by NTA. The chromatograms are shown in Figure 5.5. As shown in Fig. 5.5, the amount of canola protein bound to each column decreased in the order Cu^{2+} (68.2%), Ni^{2+} (20.6%), Zn^{2+} (9.6%), and Co^{2+} (5.7%). This suggests that, like Me^{2+}-IDA, the recognition ability of Me^{2+}-NTA for surface histidine distribution followed the order of $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$. The notable difference from the IDA results is that 5.7% of canola protein did bind to Co^{2+}-NTA compared to almost no binding to Co^{2+}-IDA. However, the separation between the canola protein and GUSH6 peaks is sufficient on Co^{2+}-NTA that SDS-PAGE (not shown) on six fractions (25 times concentrated) with the highest GUS activity did not show any visible canola protein bands. In contrast, SDS-PAGE showed that there were five or six canola protein bands with molecular weight below about 31000 Da within the GUSH6 peaks on Ni^{2+}-NTA (Fig. 5.5 b).

The relative retention of GUSH6 on Me^{2+}-NTA (Figure 5.6), was similar to that on Me^{2+}-IDA (Figure 5.3) except for Cu^{2+}. Even though GUSH6 had better affinity to Ni^{2+}-NTA than Co^{2+}-NTA (Fig. 5.6), the purification performance of the protein was better on Co^{2+}-NTA (Table 5.4), as was the case for IDA. The selection of a metal ion and a chelating

Figure 5.5. GUSH6 spiked canola extract on different metal ion charged NTA-IMAC columns. a) Cu^{2+}. b) Ni^{2+}. c) Zn^{2+}. d) Co^{2+}. Experiment parameters are the same as in Figure 5.2.


Mass balance:
Wash: ~26.5%
Peak I: 68.2%
a)
Canola protein
I
Mass balance:
Wash: ~79.2%
Peak I: 20.6%
b)
Mass balance:
Wash: ~90.4%
Peak I: 8.7%
Peak II: 0.9%
c)
Canola protein
I
II
Mass balance:
Wash: ~94.3%
Peak I: 5.7%
d)
UV
Imidazole
Activity
UV at 280 nm
GUS activity, u/ml
Imidazole, mM
Time, hours

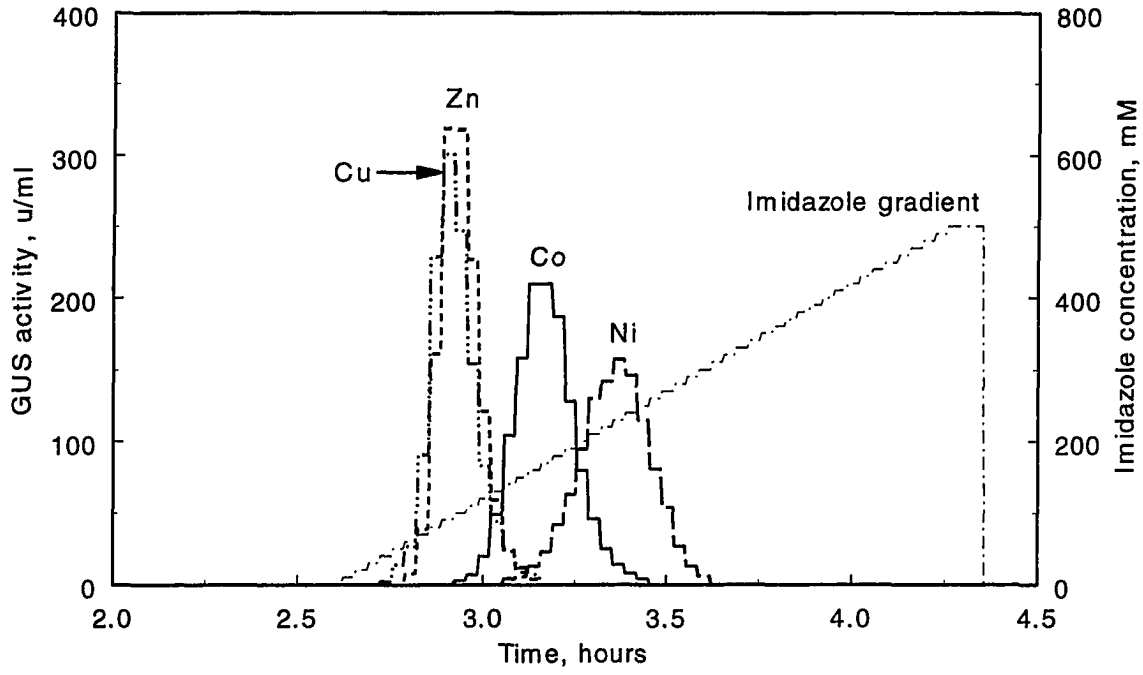


Figure 5.6. Relative retentions of GUSH6 on different Me-NTA columns. All experiments started with approximately the same total GUS activity.

Table 5.4. GUSH6 recoveries from canola extract by a single chromatographic run on various Me^{2+}-NTA columns.

Metal ions	Cu^{2+}	Ni^{2+}	Zn^{2+}	Co^{2+}
Activity before loading, U/mg	210	180	190	210
Final activity, U/mg	26,000	60,000	20,000	130,000
Recovery, %	84.9	75.2	88.8	83.3
Enrichment ratio	124	333	105	619

ligand for a his-tagged protein purification from a particular extract requires consideration of both the binding affinity of desired product relative to the impurities and the stability of chelating groups themselves (Porath, 1988b). As demonstrated in this work, Co^{2+}-IDA provided the best purification performance for GUSH6 from canola extract. However, since Me-NTA adsorbent is more stable than Me-IDA adsorbent (Hochuli, et al, 1987), Co^{2+}-NTA present a good choice when Co^{2+} leakage is the major concern.

Model of protein binding

NTA forms a tetradentate with metal ions, while IDA forms a tridentate. Simplistically then, GUSH6 should have a weaker affinity for Me-NTA's two available coordination orbitals than for the three available on the same metal ions immobilized on IDA. However, our results showed the opposite (Figure 5.7). Except for Cu^{2+}, which was irregular in several aspects, the other three metals showed stronger affinity for GUSH6 when immobilized by NTA.

Pure wild-type GUS applied to the same series of columns bound only to Cu^{2+} columns (both IDA and NTA) eluting at 70 mM and 50 mM imidazole, respectively. That is

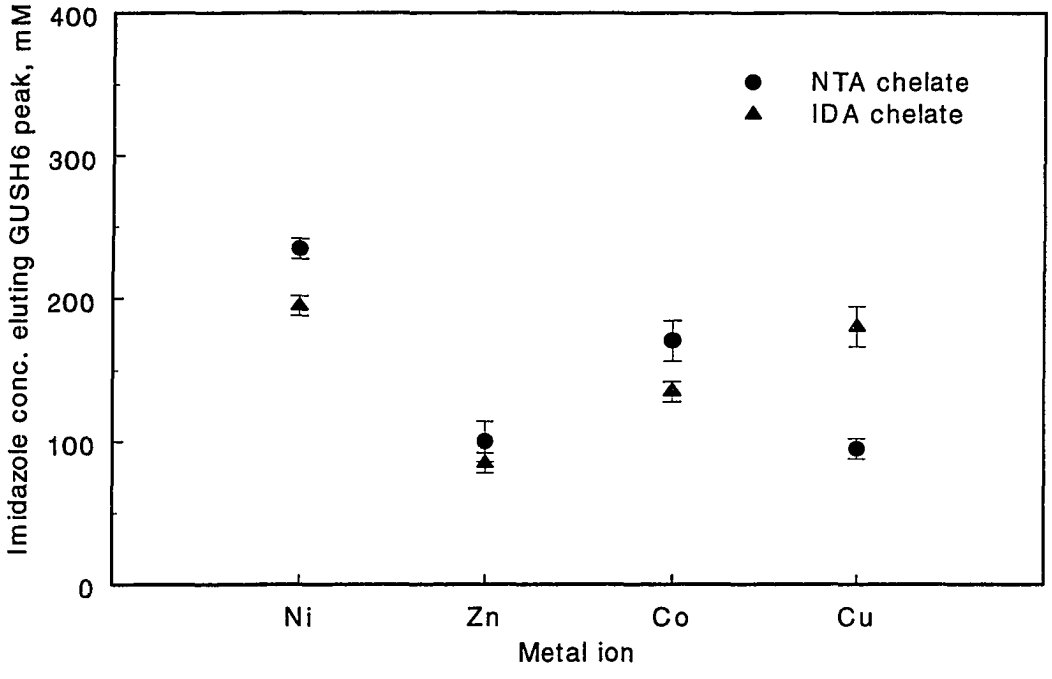


Figure 5.7. Imidazole concentrations needed to elute GUSH6 binding on different metal ions immobilized by either IDA or NTA. Errors reflect the change of imidazole concentration over each fraction time.

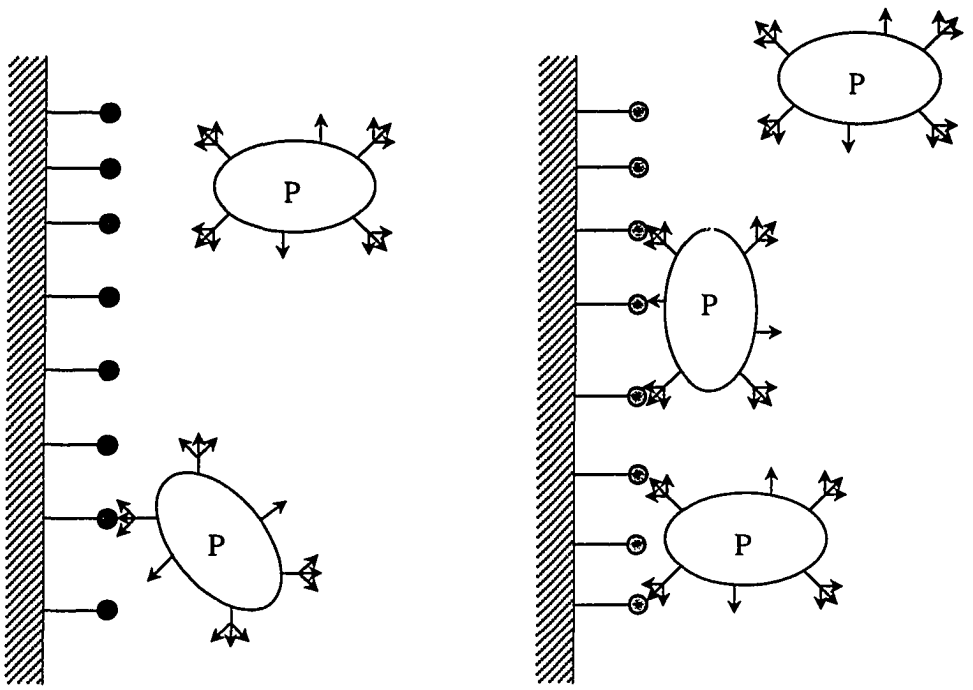
evidence that wild-type GUS has only single histidine binding sites and the multi-histidine sites are only in the fusion tail. A hypothesis as to how initial binding of the fusion tail leads to stronger binding to NTA is shown in Figure 5.8. We expect the tridentate interaction with an IDA site to be a more rigid orientation that limits additional binding by other fusions on the multimeric GUS to nearby metal sites. The greater ease of forming multi-point binding (two bidentates) with the NTA resin would then account for the higher imidazole concentration required for elution.

Protein purification from canola: IMAC or ion exchange chromatography (IEC)

Our earlier work showed that by manipulating the charge of a protein, the protein's elution could be shifted to a site with very low native canola protein background on IEC (including both cation and anion exchange), and protein enrichment factors of 50-200 could be realized in a single ion exchange step (Zhang and Glatz, 1999; Zhang, et al., 1999). Table 5.5 compares the recovery of T4 lysozyme mutants by cation-exchange and GUS mutants by anion-exchange with that of his_6-tagged GUS by IMAC. The most direct comparison is for the two GUS cases and shows IMAC to perform better.

Conclusions

This work has demonstrated IMAC could be a very effective method for purification of recombinant proteins from canola. GUSH6 enrichment factors of 600-700 can be realized by either Co^{2+}-IDA or Co^{2+}-NTA. The recognition range of metal ions for surface histidines follows the order $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$. Selection of metal ion and chelating ligand for purifying a target protein from a host extract should not solely rely on the combination giving strongest retention.



a) IDA immobilized metal ions b) NTA immobilized metal ions

Figure 5.8. Proposed protein binding mechanisms on Me^{2+}-IDA and Me^{2+}-NTA. a) On Me^{2+}-IDA; b) On Me^{2+}-NTA. The protein shown here represents a tetramer with one his-tail on each monomer (GUSH6). Tridentate formation is favored on Me^{2+}-IDA (rigid orientation), and bidentate and multi-site interaction are favored on Me^{2+}-NTA (flexible orientation).

Table 5.5. Protein recovery from canola extract by different chromatographic methods.

Method	Protein	Specific activity, U/mg	Enrichment ratio[a]
Cation-exchange (Zhang and Glatz, 1999)	Wild-type T4 lysozyme	-[b]	-
	Single mutant T4 lysozyme	385[c]	50[c]
Anion-exchange (Zhang, et al., 1999)	GUS	38,000	55
	GUSD10/15[d]	50,000/80,000	151/191
IMAC (Co^{2+}-IDA/NTA) (This work)	GUS	Unbond	Unbound
	GUSH6	141,000/130,000	705/619

a, all results were obtained by single step chromatography; for anion-exchange chromatography, total protein loading ~ 70 mg, GUS ~ 0.28 - 0.7 mg; for IMAC, total protein loading ~ 17 mg, GUS ~ 0.1 mg.

b, eluted with canola protein.

c, unpublished results.

d, D# denotes the number of aspartate in the fusion, D0 = no fusion (wild-type), D10 = 10 aspartates, D15 = 15 aspartates.

Acknowledgment

This work was supported by the NSF under Grant No. BES9522644 and the USDA under NRICGP Grant No. 95031130. The authors are very grateful to Dr. Joe Jilka of ProdiGene, College Station, TX, for providing *E. coli* strains encoded with GUSH6. We also thank Pioneer Hi-Bred International, Inc., Johnston, IA, for providing the canola seeds. We also thank Dr. Zivko Nikolov for his constructive suggestions.

References

Andersson, L. Recognition of phosphate groups by immobilized aluminum (III) ions. J. Chromatogr. 1991, 539, 327-334.

Arnold, F. H. Metal-affinity separations: a new dimension in protein processing. Bio/Technol. 1991, 9, 151-156.

Beitle, R. R.; Ataai, M. M. One-step purification of a model periplasmic protein from inclusion bodies by its fusion to an effective metal-binding peptide. Biotechnol. Prog. 1993, 9, 64-69.

Hanse, P; Lindeberg, G.; Andersson, L. Immobilized metal ion affinity chromatography of synthetic peptides. J. Chromatogr. 1992, 627, 125-135.

Hochuli, E.; Bannwarth, W.; Dobeli, H.; Gentz, R.; Stuber, D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Bio/Technol. 1988, 6, 1321-1325.

Hochuli, E.; Dobeli, H.; Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. J. Chromatogr. 1987, 411, 177-184.

Hochuli, E. and Piesecki, S. Interaction of hexahistidine fusion proteins with nitriloacetic acid-chelate Ni^{2+} ions. Methods 1992, 4, 68-72.

Jefferson, R. A.; Burgess, S. M.; Hirsh, D. β-Glucuronidase from E. coli as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 1986, 83, 8447-8451.

Jefferson, R. A.; Wilson, K. J. The GUS gene fusion system. Plant Mol. Biol. Manual 1991, B14, 1-33.

Kato, K.; Yoshida, K.; Tsukamoto, H.; Nobunaga, M.; Masuya, T.; Sawada, T. Synthesis of p-nitrophenyl β-D-glucosiduronic acid and its utilization as a substrate for the assay of β-Glucuronidase activity. Chem. Pharm. Bull. 1960, 8, 239-242.

Linderner, P.; Guth, B.; Wulfing, C.; Krebber, C.; Steipe, B.; Muller, F.; Pluckthun, A. Purification of native proteins form the cytoplasm and periplasm of *Escherichia coli* using IMAC and histidine tails: a comparison of proteins and protocols. Methods 1992, 4, 41-56.

Mantovaara, T.; Pertoft, H.; Porath, J. Purification of human serum amyloid components (SAP) by calcium affinity chromatography. Biotechnol. Appl. Biochem. 1989, 11, 564-570.

Oswald, T.; Hornbostel, G.; Rinas, U.; Anspach, F. B. Purification of $(His)_6$*Eco*RV [recombinant restriction endonuclease *Eco*RV fused to a $(His)_6$ affinity domain] by metal-chelate affinity chromatography. Biotechnol. Appl. Biochem. 1997, 25, 109-115.

Porath, J. High-performance immobilized-metal-ion-affinity chromatography of peptides and proteins. J. Chromatogr. 1988a, 443, 3-11.

Porath, J. IMAC-immobilized metal ion affinity based chromatography. Trends Anal. Chem. 1988b, 7, 254-259.

Porath, J.; Olin, B.; Granstrand, B. Immobilized-metal affinity chromatography of serum proteins on gel-immobilized group III A metal ions. Arch. Biochem. Biophys. 1983, 225, 543-547.

Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 1975, 258, 598-599.

Ritter, S.; Komenda, J.; Setlikova, E.; Setlik, I. Immobilized metal affinity chromatography for the separation of photosystems I and II from the thermophilic cyanobacterium *Synechococcus elongatus*. J. Chromatogr. 1992, 625, 21-31.

Sharma, S. K.; Evans, D. B.; Vosters, A. F.; Chattopadhyay, D.; Hoogerheide, J. G.; Campbell, C. M. Immobilized metal affinity chromatography of bacterially expressed proteins engineered to contain an alternating-histidine domain. Methods 1992, 4, 57-67.

Sulkowski, E. Purification of proteins by IMAC. Trends Biotechnol. 1985, 3, 1-7.

Sulkowski, E. The saga of IMAC and MIT. Bioassays 1989, 10, 170-175.

Sundberg, R. J. and Martin, R. B. Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. Chem. Rev. 1974, 74, 471-517.

Wilkinson, D. L.; Ma, N.-T.; Haught, C.; Harrison, R. G. Purification by immobilized metal affinity chromatography of human atrial natriuretic peptide expressed in a novel thioredoxin fusion protein. Biotechnol. Prog. 1995, 11, 265-269.

Winzerling, J. J.; Berna, P.; Porath, J. How to use immobilized metal ion affinity chromatography. Methods 1992, 4, 4-13.

Wong, J. W.; Albright, R. L.; Wang, N.-H. L. Immobilized metal ion affinity chromatography (IMAC) - Chemistry and bioseparation applications. Sep. Purif. Methods 1991, 20(1), 49-106.

Zhang, C.-M. and Glatz, C. E. Process engineering strategy for recombinant protein recovery from canola by cation exchange chromatography. Biotechnol. Prog. 1999, 15, 12-18.

Zhang, C.-M.; Love, R. T.; Jilka, . M.; Glatz, C. E. Genetic engineering strategies for purification of recombinant proteins from canola by anion exchange chromatography: an example of β-glucuronidase. Submitted to Biotechnol. Prog. 1999.

CHAPTER 6. GENERAL CONCLUSIONS

Canola was shown to have several advantages for the recovery stages when used as a recombinant protein production host. Its rather simple elution profiles on different chromatographic resins provided possibilities for effective protein purification, and genetic engineering enabled a selected protein to be manipulated for favorable recovery from canola extract.

Native canola proteins were eluted in two major peaks on linear gradient cation-exchange chromatography, and the valley between the two peaks offered a site to which to target a recombinant protein for enhanced recovery. T4 lysozyme and its mutants (both point mutation and fusion) demonstrated that charge modification can shift the elution point into such favored sites. Single mutant T4 lysozyme with a charge of +7 at neutral pH was eluted at the target site, and the recovered protein showed a high purity (90%) in a single step purification. Fusion proteins (U1 and U2) showed the same binding strength as their point mutants counterparts, but the polyarginine tail was found to be unstable in canola extract.

On anion-exchange chromatography, the native canola protein elution profile also provided several sites for targeting genetically engineered proteins. β-Glucuronidase and its fusions with polyaspartate tails showed that with added charges, the protein could be directed from one target site to another for better purification performance. GUSD10 and GUSD15 were eluted at the site with intermediate eluent salt concentration and showed lower levels of native canola protein contamination compared to GUSD0 at the first site with lower eluent salt concentration. The additional five aspartates from GUSD10 to GUSD15 did not show the same increased level of protein retention as from GUSD0 to GUSD5 and GUSD5 to

GUSD10. However, the negatively charged tails were stable in canola extract contrary to the behavior of positively charged tails. Furthermore, the comparison between experiments on transgenic canola with recombinant GUSD0 and non-transgenic canola with spiked GUSD0 showed good agreement and validated the spiking technique as a mean for investigating a protein's elution in a host protein matrix.

GUS and its fusions were characterized by the stoichiometric displacement model in isocratic elution chromatography, and the results did indicate that GUSD15 had a different trend in the change of protein specific parameters, Z and I, from other mutants. The elution of fusions on anion-exchange chromatography was shown to be predictable by both an analytical equation and a lumped dispersion model. In addition to the capability of simulating the elution peak shape, the model simulation showed slightly better ability in capturing the change of gradient slope during elution.

Different metal ions (Cu^{2+}, Ni^{2+}, Zn^{2+}, and Co^{2+}) on two chelating ligands, iminodiacetate (IDA) and nitrilotriacetate (NTA), showed different selectivities for separation of GUS-polyhistidine fusion from the native canola proteins. Most canola proteins were retained by Cu^{2+} column, but almost no canola protein bound to Co^{2+}-IDA column and only 5.7% bound to Co^{2+}-NTA. The recognition of metal ions on protein surface histidine distribution follows the order $Cu^{2+} > Ni^{2+} > Zn^{2+} > Co^{2+}$. GUSH6 was recovered with an enrichment factor of 700 by a single step Co^{2+}-IDA and 620 by Co^{2+}-NTA. Based on the experimental results on IDA and NTA columns, a binding mechanism between his-tagged proteins and immobilized metal ions was proposed. On IDA columns, the polyhistidine tail forms a tridentate with an immobilized metal ion, with a rigid orientation that would limit the ability of other surface histidines to bind simultaneously at other sites.

On the other hand, the poly-his tail forms a more flexible bidentate allowing more protein surface histidines to interact with other metal ions. Consequently, a higher imidazole concentration would be needed to elute a protein on a NTA column than an IDA column with the same metal ion.

APPENDIX A. COMPOSITION OF GUS AND FUSIONS

Amino acid composition of GUS (*E. coli*) (Jilka, personal commincation) is provided as following. It is almost identical to that reported by Jefferson, et al. (1986).

GUSD0

(N)MKYLLPTAAAGLLLLAAQPAMAMVRPVETPTREIKKLDGLWAFSLDRENCGIDQ
RWWESALQESRAIAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRF
DAVTHYGKVWVNNQEVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTI
PPGMVITDENGKKKQSYFHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCN
HASVDCQVVANGDVSVELRDADQQVVATGQGTSGTLQVVNPHLCQPGEGYLYELC
VTAKSQTECDIYPLRVGIRSVAVKGQQFLINHKPFYFTGFGRHEDADLRGKGFDNVL
MVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVVIDETAAVGFNLSLGIGF
EAGNKPKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMWSIANEPDTRPQGA
REYFAALAEATRKLDPTRPITCVNVMFCDAHTDTISDLFDVLCLNRYYGWYVQSGD
LETAEKVLEKELLAWQEKLHQPIIITEYGVDTLAGLHSMYTDMWSEEYQCAWLDM
YHRVFDRVSAVVGEQVWNFADFATSQGILRVGGNKKGIFTRDRKPKSAAFLLQKR
WTGMNFGEKPQQGGKQ(C)

GUSD5

GUSD0-DDDDD*

GUSD10

GUSD0-DDDDDDDDDD*

GUSD15

GUSD0-DDDDDDDDDDDDDDD*

APPENDIX B. ESTIMATION OF GUS AND FUSION CHARGE

The estimated net charge of GUSD0 and its fusions are calculated using Henderson-Hasselbalch equation (Heng, 1993). The numbers of amino acids in GUSD0, which will contribute to the protein net charge, their pK_a values (Stryer, 1988), and the calculated charge contribution of each amino acid at various pH are summarized in Table II-1.

Table II-1. The charge estimation of GUSD0 at various pH.

			Charges at various pH				
Amino acid	pK_a	No. in tetramer	pH=6.0	6.50	7.00	7.50	8.00
Histidine	6.50	76	57.74	38.00	18.26	6.91	2.33
Lysine	10.00	112	111.99	111.96	111.89	111.65	110.89
Arginine	12.00	116	116.00	116.00	116.00	116.00	115.99
Tyrosine	10.00	100	-0.01	-0.03	-0.10	-0.32	-0.99
Cysteine	8.50	44	-0.14	-0.44	-1.35	-4.00	-10.57
Aspartate	4.40	160	-156.08	-158.74	-159.60	-159.87	-159.96
Glutamate	4.40	156	-152.18	-154.77	-155.61	-155.88	-155.96
C-terminus	3.10	4	-3.99	-4.00	-4.00	-4.00	-4.00
N-terminus	8.00	4	-0.04	-0.12	-0.36	-0.96	-2.00
		Total Charges:	-26.71	-52.13	-74.87	-90.47	-104.27

Equation used in calculation for Tyr, Cys, Asp, Glu, C and N-termini,

$$x_i = -n_i \frac{10^{(pH-pK_{a,i})}}{1+10^{(pH-pK_{a,i})}} \quad \text{(II.1)}$$

Equation used in calculation for His, Lys, and Arg,

$$x_i = n_i \frac{1}{1+10^{(pH-pK_{a,i})}} \quad \text{(II.2)}$$

Where, x_i is the charge of an amino acid contributed to the protein, and n_i is the number of an amino acid present in the protein.

The estimated net charge of GUSD0 and its fusions at various pH are summarized in Table II-2.

Table II-2 . The charge of different GUS at various pH.

pH	GUSD0	GUSD5	GUSD10	GUSD15
6.00	-26.71	-46.22	-65.73	-85.24
6.50	-52.13	-71.98	-91.82	-111.66
7.00	-74.84	-94.82	-114.77	-134.72
7.50	-90.47	-110.46	-130.44	-150.43
8.00	-104.27	-124.27	-144.26	-164.26